UNIVERSITÉ DE SHERBROOKE
Faculté de génie
Département de génie électrique et de génie informatique

Model-based and machine learning techniques for nonlinear image reconstruction in diffuse optical tomography

Techniques basées sur des modèles et apprentissage machine pour la reconstruction d'image non-linéaire en tomographie optique diffuse

Thèse de doctorat
Specialité : Génie électrique

Seyedrohollah ETTEHADI

Jury: Jean-Marc LINA
Elijah VAN HOUTEN
Réjean FONTAINE
Yves BÉRUBÉ-LAUZIÈRE (directeur)

Sherbrooke (Québec) Canada Décembre 2017

RÉSUMÉ

La tomographie optique diffuse (TOD) est une modalité d'imagerie biomédicale 3D peu dispendieuse et non-invasive qui permet de reconstruire les propriétés optiques d'un tissu biologique. Le processus de reconstruction d'images en TOD est difficile à réaliser puisqu'il nécessite de résoudre un problème non-linéaire et mal posé. Les propriétés optiques sont calculées à partir des mesures de surface du milieu à l'étude. Dans ce projet, deux méthodes de reconstruction non-linéaire pour la TOD ont été développées. La première méthode utilise un modèle itératif, une approche encore en développement qu'on retrouve dans la littérature. L'approximation de la diffusion est le modèle utilisé pour résoudre le problème direct. Par ailleurs, la reconstruction d'image à été réalisée dans différents régimes, continu et temporel, avec des mesures intrinsèques et de fluorescence. Dans un premier temps, un algorithme de reconstruction en régime continu et utilisant des mesures multispectrales est développé pour reconstruire la concentration des chromophores qui se trouve dans différents types de tissus. Dans un second temps, un algorithme de reconstruction est développé pour calculer le temps de vie de différents marqueurs fluorescents à partir de mesures optiques dans le domaine temporel. Une approche innovatrice a été d'utiliser la totalité de l'information du signal temporel dans le but d'améliorer la reconstruction d'image. Par ailleurs, cet algorithme permettrait de distinguer plus de trois temps de vie, ce qui n'a pas encore été démontré en imagerie de fluorescence. La deuxième méthode qui a été développée utilise l'apprentissage machine et plus spécifiquement l'apprentissage profond. Un modèle d'apprentissage profond génératif est mis en place pour reconstruire la distribution de sources d'émissions de fluorescence à partir de mesures en régime continu. Il s'agit de la première utilisation d'un algorithme d'apprentissage profond appliqué à la reconstruction d'images en TOD de fluorescence. La validation de la méthode est réalisée avec une mire aux propriétés optiques connues dans laquelle sont inséres des marqueurs fluorescents. La robustesse de cette méthode est démontrée même dans les situations où le nombre de mesures est limité et en présence de bruit.

Mots-clés : tomographie optique diffuse de fluorescence, mesures optiques dans le domaine temporel, mesures optiques en régime continu, tomographie optique diffuse multispectrale, apprentissage machine.

ABSTRACT

Diffuse optical tomography (DOT) is a low cost and noninvasive 3D biomedical imaging technique to reconstruct the optical properties of biological tissues. Image reconstruction in DOT is inherently a difficult problem, because the inversion process is nonlinear and ill-posed. During DOT image reconstruction, the optical properties of the medium are recovered from the boundary measurements at the surface of the medium. In this work, two approaches are proposed for non-linear DOT image reconstruction. The first approach relies on the use of iterative model-based image reconstruction, which is still under development for DOT and that can be found in the literature. A 3D forward model is developed based on the diffusion equation, which is an approximation of the radiative transfer equation. The forward model developed can simulate light propagation in complex geometries. Additionally, the forward model is developed to deal with different types of optical data such as continuous-wave (CW) and time-domain (TD) data for both intrinsic and fluorescence signals. First, a multispectral image reconstruction algorithm is developed to reconstruct the concentration of different tissue chromophores simultaneously from a set of CW measurements at different wavelengths. A second image reconstruction algorithm is developed to reconstruct the fluorescence lifetime (FLT) of different fluorescent markers from time-domain fluorescence measurements. In this algorithm, all the information contained in full temporal curves is used along with an acceleration technique to render the algorithm of practical use. Moreover, the proposed algorithm has the potential of being able to distinguish more than 3 FLTs, which is a first in fluorescence imaging. The second approach is based on machine learning techniques, in particular deep learning models. A deep generative model is proposed to reconstruct the fluorescence distribution map from CW fluorescence measurements. It is the first time that such a model is applied for fluorescence DOT image reconstruction. The performance of the proposed algorithm is validated with an optical phantom and a fluorescent marker. The proposed algorithm recovers the fluorescence distribution even from very noisy and sparse measurements, which is a big limitation in fluorescence DOT imaging.

Keywords: fluorescence diffuse optical tomography, time-domain optical measurements, continuous-wave optical measurements, multispectral diffuse optical tomography, machine learning.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor **Prof. Yves Bérubé-Lauzière** for the great support of my Ph.D study, for his patience, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined finishing thesis without having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Réjean Fontaine, Prof. Elijah Van Houten, and Dr. Jean-Marc Lina, for their insightful comments, but also for fast reading my thesis in a delicate time.

I thank my fellow lab mates in for the stimulating discussions and for all the fun we have had in the last four years.

Last but not the least, I would like to thank my family: my beloved mother Zeynab, my father Ali, my brothers Saeed and Majid, and my lovely sister Fatemeh for supporting me spiritually throughout writing this thesis and my my life in general.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Diffuse optical tomography

1.1.1 Light interaction with biological tissues

Interaction between photons and biological tissues can be divided into two phenomena: *absorption* and *elastic scattering*. Absorption relates to light energy attenuation in tissue and elastic scattering refers to a changes in photons directions with conservation of energy. These interactions are characteristic of biological tissues and are quantified in terms of their representative coefficients: the *absorption coefficient* and the *scattering coefficient* [134]. These parameters are reconstructed in diffuse optical tomography (DOT) to identify different types of tissues in the imaged volume and for functional imaging based on tissue oxygenation. [22].

Absorption coefficient

The absorption coefficient (μ_a) quantifies the absorption of photons by a tissue. It is defined as the inverse of the average distance travelled by a photon in the tissue before it is absorbed [134]. The unit for μ_a is 1/Length (typically 1/cm or 1/mm). Biological tissues are preferably imaged with near infrared (NIR) light in the range from 650 nm to 1000 nm, so-called the "therapeutic window" as light in that range can penetrate deeper into tissues owing to much weaker absorption; several centimeters are possible, whereas in the visible range, light is strongly absorbed after only a few millimeters. The cause of light absorption in tissue can be traced back to the interaction between the different components composing tissues and photons. The main absorbers in tissues are water, fat, and both oxygenated (HbO_2) and deoxygenated (Hb) hemoglobin. The absorption spectra for these components are shown in Fig. 1.1.

Scattering coefficient

Scattering in tissues originates from inhomogeneities in the refractive index due to the presence of cell membranes and sub-cellular structures. In a similar definition to the absorption coefficient, the scattering coefficient (μ_s) is defined as the inverse of the average distance traveled by a photon before it undergoes scattering [134]. Another parameter that

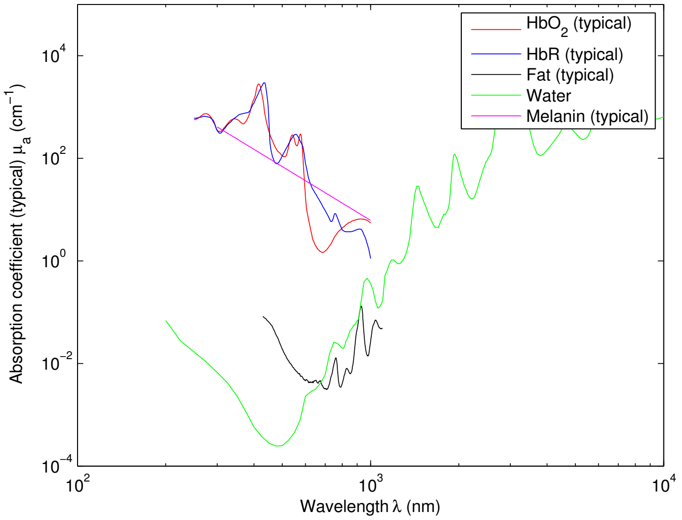


Figure 1.1 ***Absorption spectra of tissue components [132].***

must be determined in scattering is the change of direction in the propagation of a photon after it undergoes a scattering event (deflection angle). This change in direction can be characterized by the anisotropy factor (g) of the tissue. This parameter is defined as the mean cosine of the scattering angle and its typical values for biological tissues are in the range from 0.69 to 0.99 [30]. In the diffusion regime, the combination of both parameters defines the reduced scattering coefficient given by

$$\mu_s' = (1 - g)\mu_s \tag{1.1}$$

The reduced scattering coefficient appears often in the literature in relationship with the use of the diffusion equation (DE) to describe light propagation in biological tissues. The DE is a widely used approximation to the fundamental radiative transfer equation (RTE) that is the most complete macroscopic model describing light propagation in absorbing and scattering media (so-called turbid media) such as biological tissues [164]. Another optical coefficient often used, which is a combination of μ_a and μ_s' is the diffusion coefficient D given by

$$D = \frac{1}{3(\mu_a + \mu_s')}. \tag{1.2}$$

1.1.2 Definition of diffuse optical tomography and variants

Diffuse optical tomography is a biomedical optical imaging technique in which the biological tissue is illuminated (preferably by *NIR* light). Then, the multi-scattered (diffused) light exiting the medium is collected at the boundary thereof with an array of detectors arranged around the medium. The dependence on the absorption and scattering coefficients of the transmitted light through tissues allows mapping the internal optical properties of the tissue by resorting to tomographic imaging techniques. Thus, DOT for intrinsic imaging is based on specific intrinsic optical characteristics of tissue. A variant of DOT is fluorescence DOT (FDOT), in which the concentration of a fluorescent agent (typically injected into a tissue) or some other property thereof such as the fluorescence lifetime, is being imaged using the same type of measurements as for DOT. FDOT will be discussed in more depth later in this chapter. There are also other variants of DOT, such as bioluminescence DOT, but these will not be discussed here as they are not within the scope of this thesis.

The other forward models are statistical minescence.

1.1.3 Optical signal acquisition

Three main data acquisition methods in DOT are: time-domain (TD) measurements (Fig. 1.2a), frequency-domain (FD) measurements (Fig. 1.2b) and continuous-wave (CW) or steady-state measurements (Fig. 1.2c). These three methods are reviewed separately in the following.

Time-domain (TD) In TD measurements, a temporal distribution of light is produced when a short laser pulse (a few picoseconds) is transmitted through the tissue. This temporal distribution is known as the time point-spread function (TPSF). By traveling through the soft tissue, the TPSF temporally broadens and its peak intensity becomes smaller. These effects on the pulses are influenced by the optical properties of the medium and the

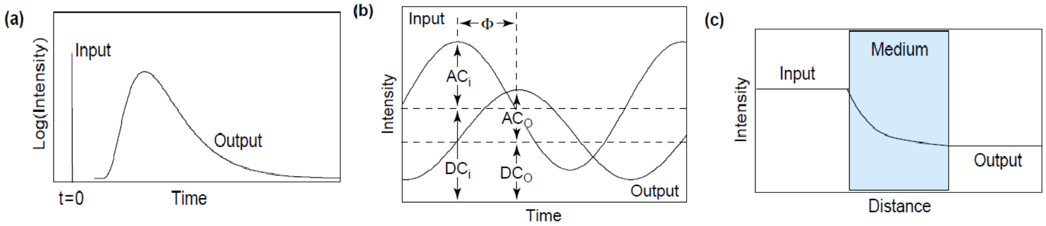


Figure 1.2 ***Optical signal acquisition methods [67].***

distance traveled therein before being a measurement is being made. In this way, a TPSF implicitly contains information about the optical properties of a medium, and methods have devised to exploit this information in view of imaging these properties. For example, the time at which the maximum of TPSF is reached is indicative of μ_s and the TPSF's tail and slope yield information about μ_a [67]. The TPSF can be temporally resolved using streak cameras or other time-gated devices, or through the use of time-correlated single-photon counting (TCSPC). Streak camera have a higher temporal resolution than TCSPC system, but the latter have a larger detection area, better temporal linearity, lower cost and higher dynamic range [74]. In TD measurements, to reach a 3D image, an array of sources and detectors around the subject are used to collect sufficient data for 3D image reconstruction. In TD tomographic imaging approaches, sources and detectors illuminate the surface of the tissue and acquire data over several views around that surface. TD methods like other methods can also use measurements at different wavelengths [58] (multispectral measurement).

Frequency-domain (FD) In frequency-domain systems, frequency-modulated light sources are used to illuminate the surface of the object. The measured parameters are the phase shift (Φ), and the demodulation ($M = (AC_o/DC_o)/(AC_i/DC_i)$)(Fig. 1.2b) [67]. A frequency-domain system consists of a radio-frequency (RF, typically 100-1000 MHz) oscillator to drive a suitable laser diode and a high-bandwidth detector (*e.g.* PMT, APD, or CCD) to detect the modulated transmitted signals [58]. The detected signals are converted from RF to a few kHz prior to phase detection. Then they are digitized over an appropriate period of time, and the phase and amplitude are computed. FD devices are relatively inexpensive and simple to use compare to TD systems. They are suitable for clinical settings because of their ability to reject non-correlated light sources such as background light. Although, FD systems are cheaper and faster to implement than TD instruments, they don't yield accurate data as TD systems [58].

continuous-wave (CW) basic optical instruments were initially developed to investigate head, breast and testes; these instruments used continuous-wave sources available at the time (Bright 1831, Curling 1843, Cutler 1929). After the introduction of NIR sources and detectors, CW systems were commonly used for studying haemodynamic and oxygenation changes in superficial tissues, and in outer (cortical) regions of the brain [58]. In CW systems, the source emits light at a constant intensity, or light that is modulated at a low frequency (a few kHz) before being injected into the tissue. Then the transmitted light intensities are measured by the detectors. CW systems, in comparison with others TD and FD data acquisition systems, reveal less information about tissues and their characteristics. It is also a difficult task to measure the absolute intensity as it requires delicate calibration and is subject to source and surface coupling fluctuations. Nevertheless, they are now among the most widely used systems in the clinical applications, owing to the availability of low-cost, high-power (several mW) laser diodes over a broad range of NIR wavelengths, simple detection instruments and high data-acquisition rates. These factors have contributed in making CW systems a popular choice for several optical imaging applications. CW-DOT has the advantages of low cost of installation and higher signal-to-noise ratio (SNR). The suitable wavelength range for optical imaging of biological tissues beyond a depths of few millimeters is between 650 to 930 nm [80], owing to the fact that absorption of the main chromophores (mainly oxy- and deoxy-hemoglobin) is least in that range.

1.1.4 DOT applications

DOT, as an emerging medical modality, covers a wide range of scales and many different applications. Since the main source of optical contrast in tissue is oxy- and deoxy-hemoglobin in blood, frequent applications of DOT in human imaging are the screening and diagnosis of breast cancer [66], and brain studies [14], including stroke, hemorrhage, and brain function [22]. Also arthritic finger imaging is well developed [68]. However, the greatest impact of DOT to date has been in small animal imaging. In molecular imaging, a wide range of biologically relevant molecules can be tagged with a broad choice of optical contrast agents. DOT imaging provides great flexibility in both the mechanisms for contrast and the measurement approaches [3].

DOT has the potential of being a low-cost imaging modality with acceptable accuracy with respect to other medical imaging techniques [2]. As mentioned, the simplest approach for implementing DOT imaging is through the use of CW light sources and light detection schemes. The major drawback of CW-DOT is its incapability of reconstructing optical properties simultaneously; more precisely, CW-DOT reconstruction is incapable of uniquely distinguishing between absorption and scattering [8]. The only solution to overcome this non-unicity is to resort to multispectral CW-DOT [12]. In multispectral tomography, optical measurements are performed at several wavelengths. This exploits, the dependency of tissue chromophore absorption and scattering to the wavelength, which allows to uniquely reconstruct the optical properties simultaneously [32, 34]. One of the recent promising applications of multispectral DOT that is currently being developed is for the detection of blood vessels during brain needle biopsy procedures [131].

1.2 FDOT

Whereas DOT seeks to image the intrinsic optical properties of tissues, as briefly mentioned previously variants of DOT have been developed to image the distribution of extrinsic compounds (i.e. that do not naturally appear in biological tissues) such as injected fluorescent probes to target specific molecules *in vivo*, and genetically engineered reporter fluorescent proteins or enzymes such as luciferase partaking in bioluminescent reactions. These approaches have been intensely developed in recent years for small animal preclinical imaging [57, 150].

1.2.1 Definition

Fluorescence is a physical phenomenon in which a molecule emits light due to excitation by an external light source. Microscopically, an excitation photon collides and is absorbed with a given probability. If absorption occurs, then the molecule transfers the photon's energy to an electron, which then leads to an excited state of the molecule. In this state, the molecule becomes unstable and tends to return to its initial state and lose energy (Fig. 1.3a). In a fluorescent molecule (fluorophore), there are two characteristic spectra, the excitation and emission spectra (Fig. 1.3b). The two diagrams are similar and the difference is in their wavelengths and/or maximum energy in excitation and emission. This difference is known as Stoke's shift, and is due to the molecular vibration of the excited fluorophore and dissipation of energy as heat before fluorescent light is emitted [108]. Thus, the energy ($E = h\nu$) of the emitted photon is smaller than that of the absorbed photon. As a result the wavelength ($\lambda = C/\nu$) of emission light is longer than the excitation light.

The intensity ratio of emitted photons to absorbed photons defines the fluorescence yield. The fluorescence yield thus gives the efficiency of the fluorescence process. This parameter depends on two properties of a fluorophore, the *extinction coefficient* and the *quantum yield* (ζ). The *extinction coefficient* is related to the absorption probability, and the *quantum yield* is the probability that a fluorescence photon will be radiated by an excited molecule. Factors such as temperature, pH and photo bleaching can affect the *quantum yield* [37, 108]. The quantum yield is one of the most important characteristics of a fluorophore. Another important characteristic is the *fluorescence lifetime* (FLT). The lifetime of a fluorophore is the average time that fluorophore remains in its excited state prior to returning to its ground state. The lifetime of molecules is typically of a few nanoseconds. The lifetime depends on the nature of the fluorophore and the chemical parameters of the local environment, like pH, oxygenation, free ion concentrations, glucose or other analytes. FLT measurements require sophisticated optics and electronics because of the short timescale of fluorescence. One of the most popular approach to measure FLT is to resort to time-domain measurements (the other being frequency-domain measurements) [59, 103, 107, 151].

Fluorescence imaging relies on an external source of light, typically a laser, to excite the fluorescent compound and then detect the emitted signal. In 1999, O'Leary reconstructed the first fluorescent molecule distribution in a turbid medium [115], and the first *in vivo* results were reported in the early 2000s [126]. This new imaging technology started to develop under the names of fluorescence molecular tomography (FMT) and fluorescence diffuse optical tomography (FDOT). Many different fluorescent probes have been developed to target conditions such as infection, apoptosis (programmed cell death) [76] and, in particular, cancer, including tumor growth, metastasis formation and gene expression [26].

Fluorescence, as a contrast mechanism in optical imaging, enhances the optical contrast between normal and diseased tissues. In FDOT, an optical filter splits the excitation signals from the external source and the emission signal from the fluorophore, based on the difference in their wavelengths. A 3D map of the fluorophore (*e.g.* of its concentration, FLT, or others) then be estimated by feeding these signals to a tomographic image reconstruction algorithm.

(a) Jablonski diagram.

(b) Stoke's shift.

Figure 1.3 ***Fluorescence: Excitation and de-excitation of a fluorescent molecule (Jablonski diagram) and Stokes' shift between the absorption and emission spectra.***

1.2.2 FDOT applications

Whole body FDOT small animal imaging has widened its applications and improved the capabilities for preclinical researches, thanks to development in photonic technology and optical biomolecular probes, such as fluorescent probes, that have led to extensive exploration of *in vivo* biological processes. Study of human disease [63, 117, 122], drug delivery and response of tissue to drugs [116] are some important applications of small animal imaging. Conventional medical tomographic imaging systems, such as PET, SPECT, CT and MRI scanners, are now commonplace in preclinical research. With rapid advances in optical imaging technologies, several optical imaging systems have been developed by labs and research centers for a wide variety of applications in preclinical and clinical research [62, 107, 114, 118, 144, 148]. Non-invasiveness, the relatiely low price of optical imaging systems and the availability of and ease of use light-emitting molecular traceable probes to target a wider range of biomolecular processes has attracted a strong interest in small animal optical molecular imaging [116]. Utilizing these probes in FDOT allows visualizing a 3D image that can provide a quantitative map of the functional and molecular status of tissue. FDOT not only can resolve the amount of a fluorescent probe that a tissue can uptake, but can also provide information about molecular biochemical binding or the pH in the local environment [124].

In small animal imaging, FDOT is of high relevance, since fluorescent tracers are commonly used in small animal studies [3, 116]. The main application of FDOT thus far is small animal imaging, but medical imaging implementations are anticipated. In biological and medical applications, proteins or other cell components can be labeled with extrinsic fluorescent dyes. The functional information is provided by locating and tracking the distribution of the target in tissue that is extremely valuable in the drug development process. So far, the technology has been successfully applied in oncology [122], studies on inflammation [63], cardiovascular diseases [117], pharmacokinetics [42], bone metabolism [101], and medical imaging applications in particular for breast cancer detection [33]. The main challenges that remain in translating preclinical small animal studies to clinical human studies are the toxicity of optical probes and associated regulations, and the thicknesses of human tissues as compared to small animal tissues (centimeters versus millimeters). The latter will most likely restrict the use of FDOT for human diagnostics to niche applications, such shallow depth imaging or imaging of relatively easily accessible tissues.

1.3 Context of the work

In this work, several 3D DOT image reconstruction algorithms have been developed for different DOT applications such as intrinsic CW multispectral and TD fluorescence DOT imaging. The algorithm for CW multispectral imaging is developed to be able to reconstruct both absorption and scattering coefficients simultaneously from multi-wavelength CW measurements, which is one of the main applications of CW imaging systems. Another algorithm is developed to reconstruct a FLT map from TD fluorescence measurements. A prototype of an FDOT scanner has been developed in our lab. The scanner uses pulsed laser excitation to acquire TD light signals at a plurality of positions around the subject without contact using detectors in a ring geometry. A primary goal of the present thesis work was to develop an image reconstruction algorithm for use with this TD optical scanner. One of the main challenges was to develop an algorithm that could work well in practice with real measurements. TD reconstruction algorithms that use full information of measured time curves were not developed before because using this full information makes the reconstruction algorithms very expensive in terms of memory and time. Here, a TD model-based iterative image reconstruction algorithm is developed that uses all acquired information of time curves to reconstruct a 3D volume of of interest inside a medium, such as a mouse, and this in a reasonable time. Moreover, a machine learning-based image reconstruction algorithm is developed for fluorescence tomography as an alternative to iterative model-based image reconstruction algorithm. The machine learning-based algorithm is able to give accurate image results even for sparse measurements. It has the capability to reconstruct a 3D image in seconds. This makes the learning-based algorithm a great tool for real-time applications in optical tomography which is a first for DOT. Also, the learning-based algorithm can extend to other medical imaging systems, since the reconstruction process is independent from the measurement strategy.

1.3.1 Motivation

Image reconstruction in (F)DOT[1] requires a precise light propagation model to predict boundary measurements (forward model). This forward model takes the form of a numerical computational model in order to be able to deal with inhomogeneous media having arbitrary geometries. In particular, in (F)DOT one must pay particular attention to the boundary of the medium, since in contrast to other imaging modalities light propagation is drastically different inside, being diffused, and outside where it is in straight lines. The forward model should be able to predict different data types such as TD, FD and CW,

---

[1] (F)DOT will be used here to denote either DOT or FDOT, or both.

as there are different methods for data acquisition in (F)DOT. The scanner developed by our group can measure time-domain optical signals, which include both CW and FD measurement. This has served as a motivation for the work in this thesis, in which image reconstruction algorithms that exploit different signal types contained in the optical signals have been developed.

CW-DOT imaging systems have the advantages of low cost of installation and higher signal-to-noise ratio (SNR). The major drawback of CW-DOT is its incapability of reconstructing optical properties simultaneously. In other words, CW-DOT reconstruction is incapable of distinguishing between absorption and scattering. A solution to overcome the non-unicity of absorption and scattering recovery and elimination of the cross-talk in CW-DOT is to resort to multispectral measurements. A model-based image reconstruction is developed to reconstruct the interior chromophores of biological tissues such as oxy- and deoxy-hemoglobin from multispectral CW measurements.

TD-DOT imaging systems provide the richest information in their signals among all the optical imaging systems, but using full TD information is computationally very expensive, owing to the need of time stepping the signals. An accelerated model-based image reconstruction is developed to overcome this difficulty and reconstruct fluorescence lifetime from TD measurements. The algorithm has the ability to provide maps distinguishing fluorophores having different lifetimes.

Model-based iterative image reconstruction algorithms are notoriously slow, owing to their iterative nature. This excludes them as a viable approach for applications that require real-time or near real-time imaging capabilities. For example, CW-(F)DOT image reconstruction for a medium of the size of a mouse takes 15-20 minutes. This time increases for TD-(F)DOT image reconstruction to a few hours. This limitation served as an incentive to look for an alternative to classical iterative reconstruction algorithms. By considering machine learning techniques and their great success in other fields and in real-time applications, the author asked the question as to whether such techniques could be applied to (F)DOT image reconstruction problems. This has led to the machine learning-based image reconstruction algorithm presented in this thesis, which delivers much faster image reconstruction suitable for real-time (F)DOT.

1.3.2 Contributions

The overall objective of this work is to develop practical image reconstruction algorithms for (F)DOT. To reach this objective, a light propagation model is first developed to predict CW and TD data types for intrinsic and fluorescence signals. First, 3D multispectral CW model-based DOT image reconstruction is developed using several optimization algorithms to determine which optimization technique is most efficient in terms of computation time, memory requirements, and image quality. This is the first time such a vast range of techniques are applied to a multispectral DOT problem. CW multispectral DOT allows reconstructing images of the distribution of chromophores inside biological tissues. The important case of oxy- and deoxy-hemoglobin imaging is considered. Next, a TD fluorescence lifetime model-based image reconstruction algorithm is developed. It allows localizing and quantifying several FLTs simultaneously. This is the first time this is achieved in FLT tomography. Finally, a completely novel machine learning-based image reconstruction algorithms is proposed and validated for FDOT. It is the first time such an algorithm is developed in (F)DOT. A key advantage of such an algorithm is its real-time imaging capability. Furthermore, this algorithm could eventually be easily adapted to other imaging modalities such as nuclear imaging (PET and SPECT), since, as will be seen, it is independent from the data acquisition strategy; the algorithm only relies on a training data set, which could easily be generated in the case of PET and SPECT in an analogous manner as done here for FDOT.

1.4 Thesis outline

A brief review of fundamental basic theories and state-of-the-art techniques in optical image reconstruction is presented in chapter 2. Details of the light propagation model, the forward and inverse problems and how to solve them are discussed. Also, a brief introduction of applied machine learning techniques (neural networks) is provided.

Chapter 3 reviews the TD (F)DOT scanner developed in our group. Most importantly, it describes the calibration technique developed for the present work in order to use for the measured signals it provides. This chapter also discusses how the calibrated measurements are processed before being fed to reconstruction algorithms.

Chapter 4 presents the model-based iterative image reconstruction algorithms using multispectral CW measurements. The details of the reconstruction algorithm and the different optimization techniques used are discussed. This chapter is about to be submitted to the Journal of Quantitative Spectroscopy and Radiative Transfer.

Chapter 5 is dedicated to the full-curve TD reconstruction algorithm for FLT tomography. The details of its implementation are discussed along with results. This chapter has been submitted as a research paper to the journal Computerized Medical Imaging and Graphics, and is under review.

The machine learning-based model for image reconstruction is discussed in chapter 6. Details of the proposed model, its implementation, and reconstructed results for a real measurement are described in detail. This chapter has been submitted to the IEEE Transactions on Medical Imaging and is under review.

CHAPTER 2

STATE OF THE ART

2.1 Light propagation model

2.1.1 Radiative transfer equation

The most general and accurate light propagation model for macroscopic absorbing and scattering media such as biological tissues is the radiative transfer equation (RTE) [89]. Radiative transfer theory was initially introduced by Schuster in 1905 (also called transport theory in other fields). It is based on the notion of energy transport in a medium containing particles. The full RTE is given by

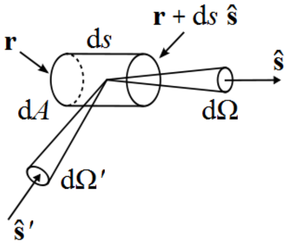


Figure 2.1 ***Geometry of light propagation.***

$$\frac{\eta}{c}\frac{\partial L\left(\mathbf{r},\hat{\mathbf{s}},t\right)}{\partial t}+\hat{\mathbf{s}}.\nabla L\left(\mathbf{r},\hat{\mathbf{s}},t\right)+\left[\mu_a(\mathbf{r})+\mu_s(\mathbf{r})\right]L\left(\mathbf{r},\hat{\mathbf{s}},t\right)= \mu_s\left(\mathbf{r}\right)\int_{4\pi}L\left(\mathbf{r},\hat{\mathbf{s}}',t\right)p\left(\hat{\mathbf{s}}',\hat{\mathbf{s}}\right)d\Omega'+q\left(\mathbf{r},\hat{\mathbf{s}},t\right), \tag{2.1}$$

where $L(\boldsymbol{r},\hat{\boldsymbol{s}},t)$ is the radiance at point $\boldsymbol{r}$ in the medium in the direction $\hat{\boldsymbol{s}}$, and at the particular time t, η is the refractive index, c is the speed of light in the vacuum, μ_a and μ_s are the absorption and scattering coefficients of the medium, $p(\hat{\boldsymbol{s}}',\hat{\boldsymbol{s}})$ is the probability that a photon from direction $\hat{\boldsymbol{s}}'$ scatters into direction $\hat{\mathbf{s}}$, and $q(\boldsymbol{r},\hat{\boldsymbol{s}},t)$ is the volume density of radiant power per unit solid angle of the source distribution within the medium (Fig. 2.1). The RTE in the frequency-domain is obtained by replacing $\frac{\partial}{\partial t}$ by $\mathrm{i}\omega$. The general boundary condition (BC) for tissue-air cases that are customary in biomedical optics is called the partial reflection boundary condition. It takes into account Fresnel reflection [164] at the boundary between the medium (tissue) and its surroundings (air). This BC is given by:

$$L\left(\mathbf{r}',\hat{\mathbf{s}},t\right)=B_T\left(\mathbf{r}',\hat{\mathbf{s}},t\right)+R_F\left(\hat{\mathbf{s}},\hat{\mathbf{n}}\right)L\left(\mathbf{r}',\hat{\mathbf{s}'},t\right),\quad \mathbf{r}'\in\partial V,\ \hat{\mathbf{s}}.\hat{\mathbf{n}}<0, \tag{2.2}$$

where B_T is the source term on the boundary for the light that is transmitted and injected into the medium, R_F is Fresnel reflection coefficient, $\hat{\mathbf{n}}$ is outward normal at the medium's boundary, ∂V denotes the boundary of the medium, and $\hat{\mathbf{s}}.\hat{\mathbf{n}} < 0$ represents directions pointing inward into the medium. R_F can be calculated as follows by providing the refractive indices inside (n_{in}) and outside (n_{out}) the medium, and the incident (θ_i), transmitted (θ_t) and critical (θ_c) angles at the boundary [164]:

$$R_F(cos_{\theta_i}) = \begin{cases} \frac{1}{2}\left(\frac{n_{in}cos\theta_i - n_{out}cos\theta_t}{n_{in}cos\theta_i + n_{out}cos\theta_t}\right)^2 + \frac{1}{2}\left(\frac{n_{out}cos\theta_i - n_{in}cos\theta_t}{n_{out}cos\theta_i + n_{in}cos\theta_t}\right)^2, & 0 \leq \theta_i \leq \theta_c \\ 1, & \theta_c \leq \theta_i \leq \pi/2. \end{cases} \tag{2.3}$$

Modeling the emission of fluorescence in the forward model involves two stages (Fig. 2.2). In first stage, an external source S^x illuminates the medium to excite fluorophores with absorption coefficient $\mu_a^{x\rightarrow m}$ at wavelength λ^x (superscript x here stands for excitation). This is the excitation stage and its corresponding light propagation equation is as follows:

$$\begin{aligned} \frac{\eta}{c}\frac{\partial L^x\left(\mathbf{r},\hat{\mathbf{s}},t\right)}{\partial t} + \hat{\mathbf{s}}.\nabla L^x\left(\mathbf{r},\hat{\mathbf{s}},t\right) + \left[\mu_s^x(\mathbf{r}) + \mu_a^{x\rightarrow m}(\mathbf{r})\right] L^x\left(\mathbf{r},\hat{\mathbf{s}},t\right) = \\ \mu_s^x\left(\mathbf{r}\right)\int_{4\pi} L^x\left(\mathbf{r},\hat{\mathbf{s}}',t\right) p\left(\hat{\mathbf{s}}',\hat{\mathbf{s}}\right) d\Omega' + S^x\left(\mathbf{r},\hat{\mathbf{s}},t\right). \end{aligned} \tag{2.4}$$

In the second stage, the emission of fluorescence light at wavelength λ^m is calculated from fluorescence source S^m that depends on the coefficient $\mu_a^{x\rightarrow m}$ combined with the quantum yield ζ (superscript m here stands for emission). This is the emission stage and the corresponding equation is the following:

$$\begin{aligned} \frac{\eta}{c}\frac{\partial L^m\left(\mathbf{r},\hat{\mathbf{s}},t\right)}{\partial t} + \hat{\mathbf{s}}.\nabla L^m\left(\mathbf{r},\hat{\mathbf{s}},t\right) + \left[\mu_a^m(\mathbf{r}) + \mu_s^m(\mathbf{r})\right] L^m\left(\mathbf{r},\hat{\mathbf{s}},t\right) = \\ \mu_s^m\left(\mathbf{r}\right)\int_{4\pi} L^m\left(\mathbf{r},\hat{\mathbf{s}}',t\right) p\left(\hat{\mathbf{s}}',\hat{\mathbf{s}}\right) d\Omega' + \frac{1}{4\pi}\zeta\mu_a^{fl}(\mathbf{r})\psi^x\left(\mathbf{r},t\right). \end{aligned} \tag{2.5}$$

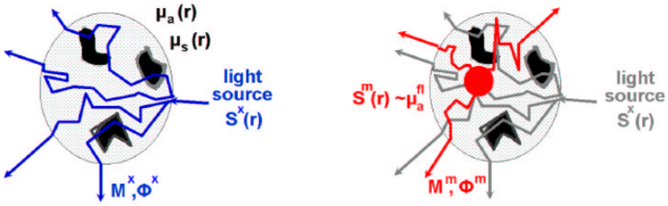


Figure 2.2 ***Fluorescent emission [92].***

Light fluence rates $\psi^{x,m}$ are obtained from the radiance $L^{x,m}$ using the following relation

$$\psi^{x,m}\left(\mathbf{r},t\right)=\int_{4\pi}L^{x,m}\left(\mathbf{r},\hat{\mathbf{s}},t\right)d\Omega. \tag{2.6}$$

The main disadvantage of the RTE is the complicated implementation within a numerical setting. The major challenge in this case resides in using an appropriate method to incorporate the angular dependence of the problem. Several approximations to the RTE have been proposed such as the diffusion approximation [164], the P_N approximation (series of spherical harmonics) [23], and more recently simplified spherical harmonics approximation [79, 97]. In this work, the diffusion approximation is resorted to for its simplicity.[1]

2.1.2 Diffusion approximation

In many practical situations in which often $\mu_s \gg \mu_a$, the simpler diffusion approximation (DA) can be used. The DA is based on two assumptions: i) the radiance is almost isotropic and ii) $\mu_s \gg \mu_a$. The second assumption means that the fractional rate of change of the current density vector ($J=\int_{4\pi}\hat{s}Ld\Omega$) times the transport mean free time shall be much smaller than unity [164]. The DA leads to the following partial differential equation, so-called the diffusion equation (DE):

$$\frac{\eta}{c}\frac{\partial\psi\left(\mathbf{r},t\right)}{\partial t}+\mu_a\left(\mathbf{r}\right)\psi\left(\mathbf{r},t\right)-\nabla\left[D\left(\mathbf{r}\right)\nabla\psi\left(\mathbf{r},t\right)\right]=Q\left(\mathbf{r},t\right), \tag{2.7}$$

where

$$D\left(\mathbf{r}\right)=\frac{1}{3\left(\mu_a\left(\mathbf{r}\right)+\mu_s'\left(\mathbf{r}\right)\right)}, \tag{2.8}$$

is the diffusion coefficient, and

$$\psi\left(\mathbf{r},t\right)=\int_{4\pi}L\left(\mathbf{r},\hat{\mathbf{s}},t\right)d\Omega \tag{2.9}$$

$$Q\left(\mathbf{r},t\right)=\mu_s\left(\mathbf{r}\right)\int_{4\pi}L\left(\mathbf{r},\hat{\mathbf{s}}',t\right)p\left(\hat{\mathbf{s}}',\hat{\mathbf{s}}\right)d\Omega' \tag{2.10}$$

---

[1] Note, however, that the algorithms presented in this thesis could have resorted to the other more sophisticated models alluded to above. This would have increased the computational burden associated with the algorithms, but fundamentally the rationale would remain the same.

are the fluence rate and isotropic source term, respectively. Associated the boundary conditions associated with the DE can be obtained from Eq.(2.2) using the aforementioned assumptions, leading to:

$$\left(\frac{1-R_\psi}{4}\right)\psi\left(\mathbf{r}',t\right)+\left(\frac{1+R_J}{2}\right)D(\mathbf{r}')\frac{\psi\left(\mathbf{r}',t\right)}{\hat{\mathbf{n}}}=b_T\left(\mathbf{r}',t\right), \tag{2.11}$$

where

$$R_\psi = 2\int_0^{\pi/2}\cos\theta_i R_F(\cos\theta_i)\sin\theta_i d\theta_i, \tag{2.12}$$

$$R_J = 3\int_0^{\pi/2}\cos^2\theta_i R_F(\cos\theta_i)\sin\theta_i d\theta_i, \tag{2.13}$$

and

$$b_T\left(\mathbf{r}',t\right)=\int_{\hat{\mathbf{s}}.\hat{\mathbf{n}}<0} B_T\left(\mathbf{r}',\widehat{\mathbf{s},t}\right)\hat{\mathbf{s}}.(\widehat{-\mathbf{n}})d\Omega. \tag{2.14}$$

where B_T is any source term located at the boundary [20]. Similarly to Eqs. (2.4), (2.5), coupled DEs can be written for a fluorescence field as follows:

$$\begin{aligned}\frac{1}{\nu}\frac{\partial\psi_x}{\partial t}+\left(\mu_a^x(\mathbf{r})+\mu_a^{x\to m}(\mathbf{r})\right)\psi_x(\mathbf{r},t)-\nabla\cdot\left[D^x(\mathbf{r})\nabla\psi_x(\mathbf{r},t)\right]=Q_x(\mathbf{r},t), \quad \mathbf{r}\in\Omega,\\ \frac{1-R_\psi}{4}\psi_x(\mathbf{r},t)+\frac{1+R_J}{2}D^x(\mathbf{r})\frac{\partial\psi_x}{\partial\hat{\mathbf{n}}}=0, \quad \mathbf{r}\in\partial\Omega,\end{aligned} \tag{2.15a}$$

$$\begin{aligned}\frac{1}{\nu}\frac{\partial\psi_x}{\partial t}+\mu_a^m(\mathbf{r})\psi_m(\mathbf{r},t)-\nabla\cdot\left[D^m(\mathbf{r})\nabla\psi_m(\mathbf{r},t)\right]=Q_m(\mathbf{r},t), \quad \mathbf{r}\in\Omega,\\ Q_m(\mathbf{r},t)=\frac{\eta}{\tau(\mathbf{r})}\mu_a^{x\to m}(\mathbf{r})\int_0^t\psi_x(\mathbf{r},t')\exp\left(-\frac{t-t'}{\tau(\mathbf{r})}\right)dt',\\ \frac{1-R_\psi}{4}\psi_m(\mathbf{r},t)+\frac{1+R_J}{2}D^m(\mathbf{r})\frac{\partial\psi_m}{\partial\hat{\mathbf{n}}}=0, \quad \mathbf{r}\in\partial\Omega.\end{aligned} \tag{2.15b}$$

where $\eta\mu_a^{x\to m}$ and τ are respectively the fluorophore's yield and FLT, both of which may vary spatially. Other forward models are statistical models based on Monte Carlo techniques or random walk theories. These forward models are commonly used in optical imaging where the DA does not hold [58].

2.2 Model-based image reconstruction

The standard back-projection technique, as applied in X-ray tomography, has been of limited success in optical imaging because of the strong scattering of light in tissues [89]. The majority of current image reconstruction techniques in (F)DOT can be categorized as model-based image reconstruction (MBIR) algorithms [6, 90]. These iterative algorithms can be divided into two steps. The first step, called the *forward problem*, consists solving light propagation model, assuming the optical properties to be known, and given the source locations to predict the measurements at the detector locations. These predictions are then compared with experimental measurements using a measure of their discrepancy through a so-called objective function. In the next step, called *inverse problem*, the medium optical properties are updated for the forward problem by minimizing the error (objective function) between the model predictions and the experimental measurements (objective function) optimization method [90]. The whole image reconstruction process consists of the forward problem and inverse problem repeated iteratively until the objective function reaches a minimum determined via certain criteria (Fig. 2.3). The final distribution of optical properties is displayed as an image.

To apply the MBIR method in the case of fluorescence tomography, the forward problem needs to be solved at two wavelengths(excitation and emission wavelengths). The emission problem is dependent on the solution of the equation in the excitation problem to provide the emission source term. The location and concentration of the fluorescent compound are initially unknown, and the reconstruction process serves to determine them. For this, the fact the total fluorophore concentration (C) is proportional to fluorophore absorption ($C \propto \mu_a^{x \to m}$) (Eq. (5.1a)) is used [92].

Depending on how the forward problem is solved, image reconstruction algorithms can be divided into two groups: linear and nonlinear image reconstruction. Light propagation inside the medium is a nonlinear problem with respect to the scattering and absorption values. Since solving the nonlinear forward problem is complex and time consuming, a linearized version of the forward model can be written using the popular Born or the Rytov approximations [55, 112, 115, 146, 168]. Linearized models are less accurate since they assume that the optical properties distribution inside the medium are a small spatial perturbations over a constant background, which is generally not true for biological tissues as there can be marked contrast in these properties between different organs. In this work nonlinear forward models are used for both DOT and FDOT.

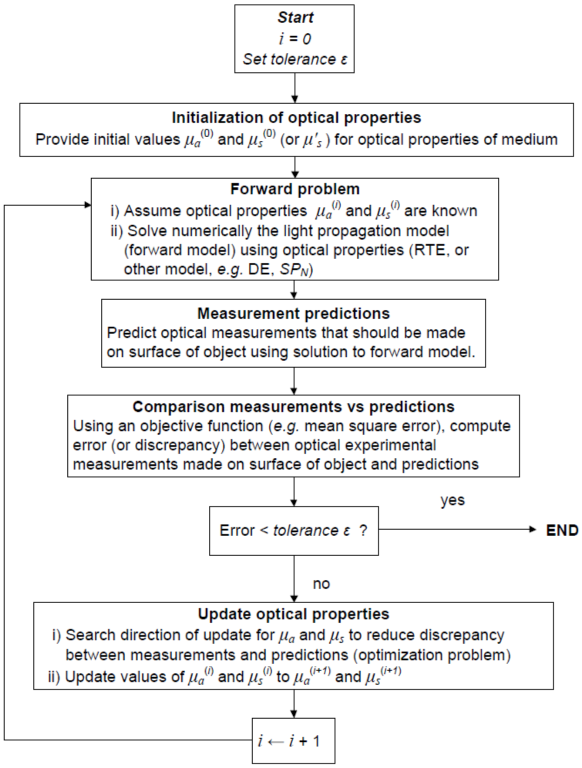


Figure 2.3 ***Reconstruction algorithm in iterative technique.***

2.2.1 Forward problem

In the forward problem, the spatial distribution of the light field is predicted inside and at the boundary of the medium given the optical properties and the sources. This predicted field is used to calculate the values of measurable quantities at the boundary such as the amount of light exiting the medium. These predicted boundary values with real measured values are used to establish an objective function.

The forward problem can be solved with an analytical or statistical or numerical model. Analytical models are difficult, if not impossible, to be solved for complex geometries and inhomogeneous media as encountered in practical cases [38, 48]. As regards statistical models, these are typically based on the Monte Carlo (MC) method [163]. These models are considered as *gold standards* against which other models should be compared [45, 48]. However, stochastic methods are computationally very demanding and solving them for practical applications is very time consuming. Furthermore, stochastic methods are hardly integrable in image reconstruction algorithms, because they do not allow any analytical equations on which such algorithms are ultimately based. Hence, numerical models must ultimately be resorted to. In these models the medium is discretized spatially, and the light propagation equation in the form of PDEs, leads to a set of discretized equations that can be solved numerically with linear algebra techniques. In this way predicted boundary measurements are calculated.

Numerical space discretization

Three general techniques for discretizing a 3D volume are: the finite volume method (FVM) [119], the finite element method (FEM) [31, 156], and the finite difference method (FDM) [93, 155]. Here, the forward model is discretized with the FDM [69, 93, 155]. The FDM is easy to implement and runs faster that the FVM and the FEM because of the structured grid underlying it. It leads to simpler banded matrices with faster algorithms to solve linear system of equations involving such matrices.

Finite Difference Method (FDM) FDM is a numerical method to solve PDEs with certain boundary conditions [157]. It approximates the PDEs' derivatives with differences on a structure grid [139]. Depending on the medium's dimensions (2D/3D), these differences can apply on a square (difference over 4 adjacent nodes) or cube (difference over 6 adjacent nodes). All the variables (spatial and/or temporal) in PDEs can be expressed with finite differences on a structure grid. There are several techniques to approximate the derivative for spatial/temporal variables. For the spatial discretization, forward, backward, or central differences are used and for the temporal discretization, an explicit (forward Euler) or, implicit schemes (backward Euler), (backward Euler or full-implicit, Crank-Nicholson, . . .) can be used. In the case of parabolic PDEs as considered here, the Crank-Nicholson scheme is favored as it is unconditionally stable.

2.2.2 Inverse problem

To reconstruct an image in (F)DOT, it is necessary to solve the inverse problem. The goal of the inverse model is the determination of unknown optical properties (absorption and scattering coefficients in DOT, and fluorescence yield and FLT in FDOT) using the measured boundary data [41]. When absolute values of the optical properties are desired or the difference between two similar states cannot be assumed small, a full non-linear reconstruction problem must be solved. To solve the non-linear problem, an objective function (OF) must be defined. The OF measures the discrepancy between real boundary measurements and the predicted values from the solution to the forward problem (Eq. (2.16)). A least square objective function (L_2 norm) in the (F)DOT can be defined as follows:

$$f(x) = \frac{1}{2} \left\| P.F(x) - y \right\|_2^2 \tag{2.16}$$

where y represents the real boundary measurements, P is the operator to extract the predicted boundary values from the solution to the forward model, and thereof F is the forward model operator that depends on the optical properties x to be reconstructed. Least square functions are commonly used as OFs since they are easy to implement and minimize. To minimize the OF and update the optical properties entering the forward model, optimization techniques are used.

2.2.3 Optimization method

Optimization to minimize a function such as the objective function in (F)DOT image reconstruction can be categorized into different classes. In a first class, one finds first derivative optimization methods [12], in a second class are quasi-Newton optimization methods [70], and in a third class on finds Newton optimization methods [6, 21, 41, 78]. These three well-known optimization classes will now be briefly described.

First derivative class

Optimization methods in this class require the function value f and its first derivative (g) to find a descent direction. The steepest descent method, conjugate gradient method, and nonlinear conjugate gradient method are some of optimization methods in this class. In all these methods a descent direction is related to the direction of $-g$.

Quasi-Newton class

Quasi-Newton methods, similarly to first derivative methods, use first derivative information only, but compute second derivative information. In fact, an approximated Hessian matrix is calculated based on a sequence of function and gradient values from previous iterations. Most of these methods force the approximate Hessian to be symmetrical and positive definite which greatly improves their convergence properties. The Davidon-Fletcher-Powell (DFP) method, Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, and its low-memory version the L-BFGS, are optimization methods in this class.

Newtons class

Newton algorithms require the function value f, and the first and second derivative (gradient g and Hessian H) to find descent direction. Newton methods use a second-order Taylor series expansion of the function about the current point in the iterations. In Newton methods, computed descent directions are along the direction of $H^{-1}g$. Scaling the direction with H in Newton methods improves the optimization performance, but at the cost of computing the Hessian matrix H and its inverse. Also, Newton methods can deal with non-normalized data, because of the scaled direction. The calculation of H matrix is very time/memory consuming and almost impractical for big problems such as time-domain (F)DOT image reconstruction. More details of optimization algorithms and their implementations are described in chapter 4.

2.3 Machine learning-based image reconstruction

Today, machine learning techniques are applied in many fields from science (physics, statistics, engineering, *etc.*) to modern society (web searches, social networks, e-commerce websites, *etc.*). Conventional machine learning techniques such as support vector machines, random forests, and neural networks with a few layers, were limited in processing real data to learn their complex functionality. In the past, building an efficient machine learning system was required careful engineering and considerable domain expertise. Designing a feature extractor that transformed the raw data into a suitable hidden representation or feature vector was very difficult and required lots of computation. For years, this made the learning process not suitable for real applications. In the last few years, by introducing new deep neural networks, it has become possible to apply machine learning techniques to a broad range of real applications. Here, a brief introduction will first be given to simple and deep neural networks. Then how to use neural networks in the problem of interest here, that is (F)DOT image reconstruction, will be discussed.

2.3.1 Neural networks

A neural network (NN) consists of several processing nodes (called neurons), arranged in several layers which are inter-connected (inspired by the structure of the brain, hence, the name neural network). Each neuron produces a real value (also called an activation value). Neurons of the first layer (input layer) are activated through the input values (Fig. 2.4).

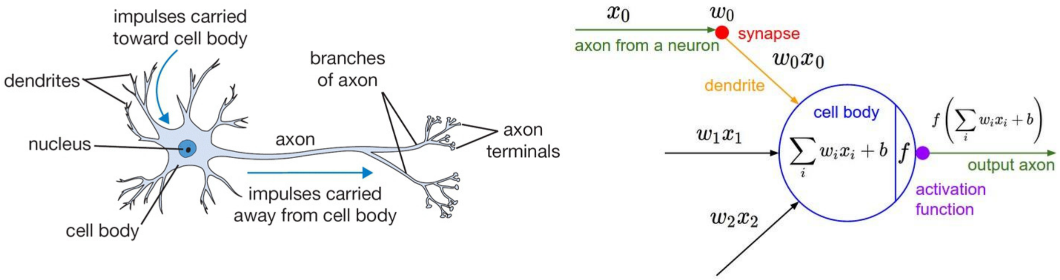


Figure 2.4 ***A cartoon drawing of a biological neuron (left) and its mathematical model (right) [84].***

The next layers (also called hidden layers) in the network are activated through weighted connections from previous layers (active neurons propagate in a forward direction). The learning process is about finding neuron weights in a NN which can demonstrate the desired functionality, such as understanding numbers. Such functionality requires a sequence of computational layers, where each layer transforms the aggregate activation of the network in a nonlinear way. The architecture of the network can be changed depending on the input/output of the problem and how the layers are connected. Increasing the number of consecutive layers enables the network to learn better nonlinear functionalities. Deep learning is about accurately learning across many consecutive layers (Fig. 2.5).

2.3.2 Deep neural networks

Deep neural network models made a breakthrough in the last few years. They attract a lot of attention in many applications. In general, deep learning is a data analysis model with several layers that promote a higher level of abstraction and prediction from data. In the training process, a large set of data is fed to the input and output layers of the network and weights of the neurons are updated along the gradient descent direction of a loss function. The important factors that enable these methods to perform exceptionally well are a bulk set of data (big data), availability of high-performance computing power, smarter weight initialization, and significantly deeper network architecture. Neural network models are often organized into distinct layers of neurons. One of the main reasons that neural networks are arranged into layers is that this architecture makes them very simple and efficient to handle using matrix vector operations. Depending how these layers are connected to each other, several architectures are introduced. Here, two main architectures that are of interest for deep models in image reconstruction are briefly presented.

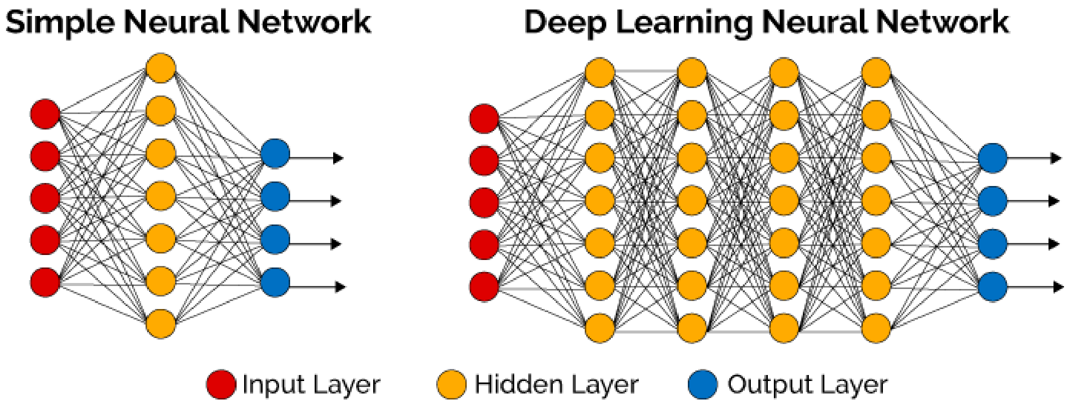


Figure 2.5 ***Simple and deep neural networks.***

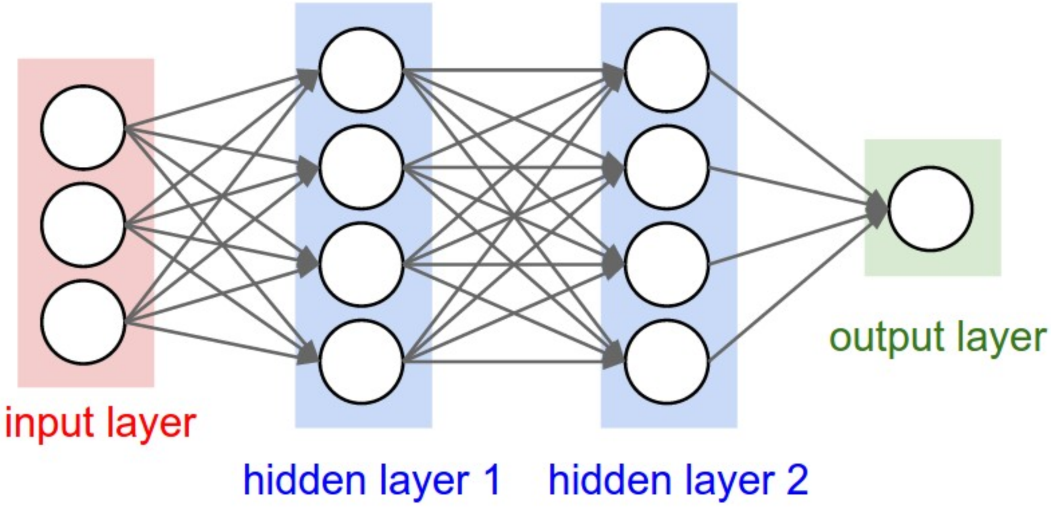


Figure 2.6 ***A full-connected layer with three inputs, two hidden layers of 4 neurons each and one neuron as output layer.***

Fully-connected layer The regular architecture in NN is the fully-connected layer (also called dense layer). In this architecture, there is no connection between neurons in the same layer, but neurons of two adjacent layers are fully pairwise connected (Fig. 2.6). All hidden layers in a NN have an activation function (also called a linear identity activation function). Several activation functions are used and introduced in NN such as the sigmoid function ($\frac{1}{(1+\exp(-x))}$), the tanh function ($\tanh(x)$), the Rectified Linear Unit (ReLU $\max(0, x)$), and the Leaky ReLU ($(x < 0)(\alpha x) + (x >= 0)(x)$ where α is a small constant). It is worth mentioning here, that in practice it is often the case that neural networks with 2 hidden layers will outperform networks with 1 hidden layer, but in fully-connected architectures going even deeper (3, 4, 5 hidden layers) rarely improves performances [84].

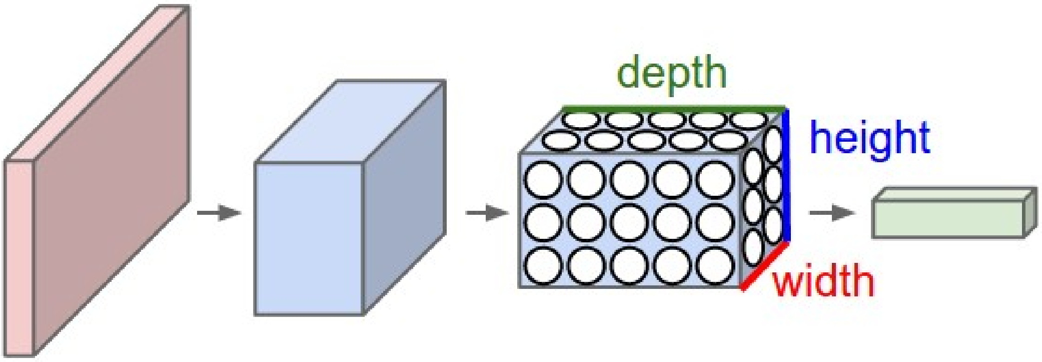


Figure 2.7 ***A Convolutional network with two hidden layers. Each layer consists of a width, height and depth, as visualized in one of the layers.***

Convolutional layer The convolutional layer architecture, similar to the fully-connected layer architecture, is made up of neurons that have weights and biases, with a difference. In the convolutional layer there is a partial connection between two adjacent layers. A filters with certain sizes (*e.g.* a 5 × 5 filter) is used in the convolutional layer to map a region of image (size of filter) to a neuron (Fig. 2.7). In fact, this architecture led to map the input, which is very often an image, to a neuron in the next layer. Here, in contrast to the fully-connected layer, more layers (deeper architecture) have been found to be an extremely important component for a good recognition system [84]. One reason for this behavior is that images consist of hierarchical structures (*e.g.* faces are made up of eyes, which are made up of edges, *etc.*), so more layers of processing make intuitive sense for this data domain.

In the last few years, one the most successful approaches that has emerged for image analysis in deep learning, is that of probabilistic generative models. These models are capable of generating high-dimensional data such as images from sampling of a low-dimensional space. Generative models and specifically conditional variants of them are a very promising choice for medical image reconstruction. A conditional generative model can reconstruct an image from given measurements (measurement distribution are considered here as condition distribution). Here, two well-known generative models are reviewed for image generation: generative adversarial networks and variational autoencoders.

2.3.3 Deep generative network

Generative adversarial networks

Generative adversarial networks (GANs) were introduced by Goodfellow *et al*, [61] to interpret images as samples from a high-dimensional probability distribution. A GAN consists of two networks: generative and discriminative. The generative network generates images

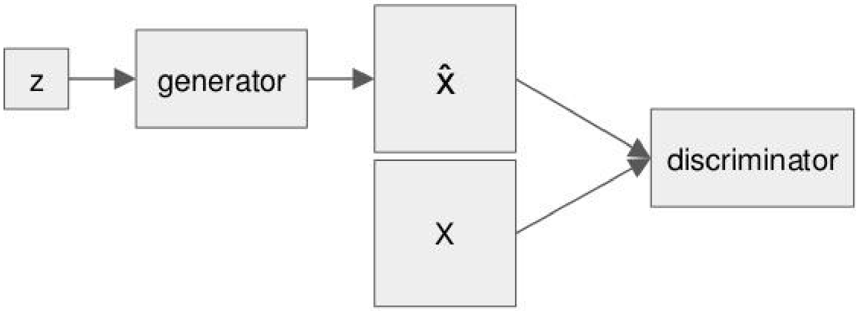


Figure 2.8 ***Scheme for generative adversarial networks.***

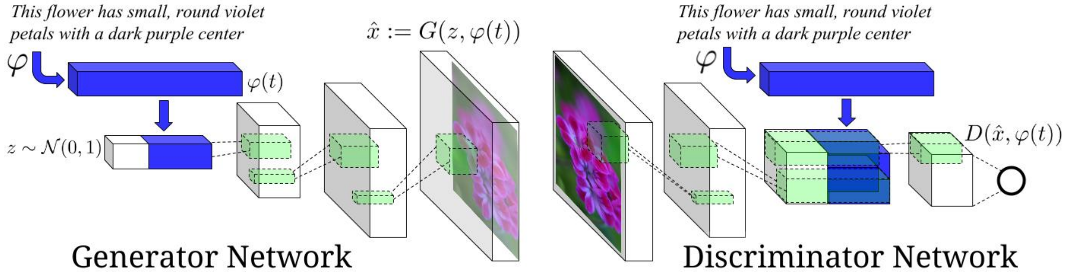


Figure 2.9 ***Schematic of a conditional GAN for image generation with given text [136].***

from a sample of a distribution (commonly Gaussian) and a discriminative network classifies the images as *real* or *fake*. Generator network is optimized to maximize the probability of the generated image (fake) close to the real image, and the discriminator is optimized to discriminate the *fake* from the *real* images (Fig. 2.8). In fact, they train each other in an adversarial way. The loss function defined in this model is a binary cross entropy with respect to the output of the discriminator. Training adversarial networks is not simple and optimization is somewhat unstable, but the results are sharper in comparison with VAE. For more details refer to [61].

Conditional generative adversarial networks are a variant of GANs which can construct data from given condition. Condition here means any extra information that is related to the observed data. For example, in Fig. 2.9, the conditional GAN can reconstruct an image by a given caption of text. The model is trained with a dataset of images and their related captions (the caption dataset is considered here as a conditional distribution of the image distribution). Then, the conditional GAN model can reconstruct an image from given dataset.

Variational autoencoder

One of the well-known deep generative models is the variational autoencoder (VAE). Autoencoders are networks with the same input/output size. First, an encoder network infers a low-dimensional dense representation of the set of inputs (simply called the input)

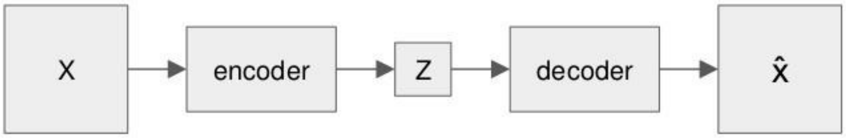


Figure 2.10 ***A scheme of variational autoencoder.***

by way of so-called latent variables (2.10). For this reason, an encoder network is also called an inference network. The output of the encoder can be interpreted as a manifold of inputs (the input manifold). Then a decoder reconstructs the set of outputs (simply called the output) from the latent variables. In variational autoencoders, instead of encoding the input to an explicit vector in the latent space, a density function (normally unit Gaussian) of each set of inputs is encoded in the latent space [88]. This encourages the model to place a high probability mass on many latent variables that could have generated samples, rather than collapsing to a single point estimate of the most likely value [60]. This Gaussian distribution makes the model less sensitive to small variations around the input such as noise or other measurement errors and turns it to a probabilistic generative model [43]. These models are simple to implement, and can be optimized with an element-wise error between input and output, *e.g.* square error (serving as loss function). This quadratic loss function is easy to optimize and commonly tends to generate blurry images [111].

Just like conditional GANs, a conditional VAE can be defined which can generate data with a given condition. An example of a conditional VAE for face generation with given conditions is depicted in Fig. 2.11. In this example, images can be modified with given states (conditions) such as pose, light, shape, *etc.*, which are added directly to the latent space.

2.3.4 Deep learning in medical imaging

The performance of deep networks is now well reported in the areas of computer vision [49], speech recognition [72], and text/image processing [36, 169]. In medical imaging analysis/processing, machine learning approaches, and more recently deep learning ap-

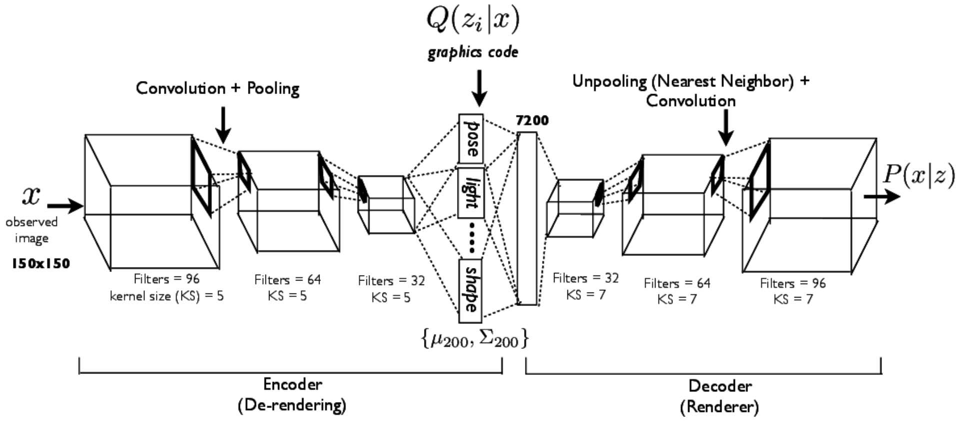


Figure 2.11 ***Schematic of a conditional VAE for image generation with given condition [102].***

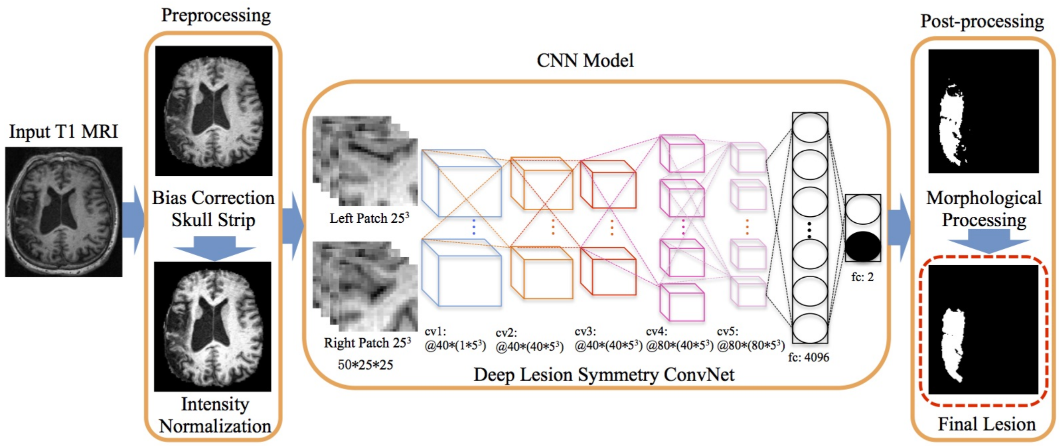


Figure 2.12 ***Overview of a deep convolutional network for stroke lesion segmentation on MRI images.***

proaches, have been applied for lesion detection [147, 149, 160], segmentation, and classification [16, 27, 56, 130, 140]. An example of a deep convolutional network for lesion segmentation is depicted in Fig. 2.12. In the medical image reconstruction field, a few approaches have been developed for de-noising low-dose X-ray CT images [83], transform from k-space to image space in MRI [171], and enhance photo-acoustic reconstructed images [5]. Here an algorithm to directly reconstruct an image from the power of received fluorescence signals will be proposed. This algorithms, as will be seen, can lead to real-time image reconstruction from noisy and sparse measurements.

In optical image reconstruction, the optical properties will be reconstructed from a set of measurements on the boundary of the medium. It means the encoded information in the measurement domain (light propagation inside medium) should be decoded to a spatial map (image). One option to implement such models can be borrowed from statistics and inference theory. In statistical image reconstruction one tries to maximize the probability of $p(y \in Y|x \in X)$, where x is a sample from the measurement space X and y is a sample from the image space Y. Inference from such probabilistic models is often intractable and a neural network model that learns approximate solutions to the inference problem can be trained. One approach to approximate inference is through variational inference algorithms. Such algorithms cast inference as an optimization problem. In this project, a deep generative model based on a VAE is developed to replace sampling from high-dimensional space $p(y|x)$, by sampling from a low-dimensional joint manifold space $z \in \mathcal{N}(0, I)$.

After reviewing state-of-the-art methods for image reconstruction in DOT imaging, data preprocessing and calibration techniques that are required to be applied on the measurement from our optical scanner will be presented at next chapter.

CHAPTER 3

OPTICAL SCANNER AND DATA PREPROCESSING

3.1 Scanner

Our group developed a non-contact time-domain DOT scanner (Fig. 3.1) with multi-view detection (over 360°) for localizing fluorescent markers in scattering and absorbing media, in particular small animals [110]. The system consists of a Ti:Sapphire ultra-short pulse laser (Mai Tai, Newport-Spectra Physics, $\approx$ 3 ps pulse width, 80 MHz repetition rate, 705 nm to 985 nm tuning range), 14 free-space optics non-contact detection channels around the medium to detect intrinsic and fluorescence signals (seven channels each), 14 photomultiplier tube (PMT) detectors to measure the optical signals at both the excitation and fluorescence wavelengths, and a time-correlated single photon counting (TCSPC) instrumentation card (SPC-134, Becker and Hickl, Berlin, Germany) to acquire the time point-spread functions (TPSFs). The free-space optical channel allows minimizing the temporal dispersion of optical signals and allows sharper IRFs in distinction to other systems resorting to fiber optics. Thanks to its free-space optics design and short laser pulses used, the scanner provides for higher timing resolution compared to other multi-view time-domain scanners. In this chapter, calibration methods and data preprocessing

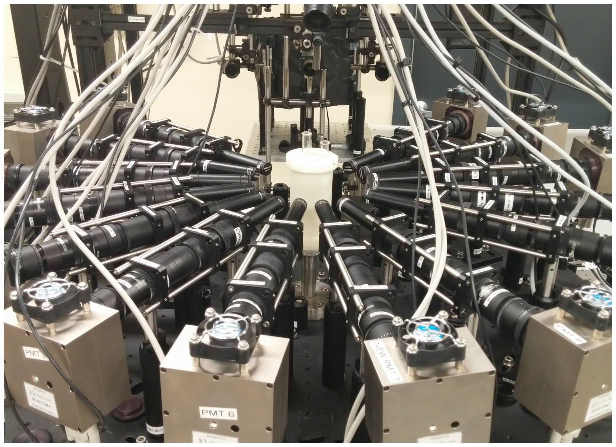


Figure 3.1 ***(F)DOT scanner.***

steps that are required to be able to use the signals measured with the scanner in the image reconstruction algorithms will be discussed.

3.2 Calibration

TD-DOT imaging systems, similarly to CW optical systems, are required to be calibrated with respect to signal intensity so that differences between different detection channels can be compensated for. Moreover, they are required to be calibrated with respect to time delays in the signal paths (both optical and electronics delays) and the temporal dispersion of the signal. Calibrating the optical scanner with respect to all these factors is difficult. Here, the image reconstruction algorithm is developed to compensate part of these calibration difficulties. For this purpose, an energy normalization is applied to TPSFs before feeding them to the reconstruction algorithm. Then, only a relative calibration is necessary. In relative calibration, channels are calibrated with respect to one reference channel relatively. This relative calibration corrects the TPSFs for the differences in detector sensitivity, gain, and time delay as well as any differences in filter performance in each channel.

3.2.1 IRF measurement

In TD systems, an optical signal emanating from the object to be imaged (which will be called a *true* signal as it has not been altered by the measurement system) is temporally convolved with the instrument response function (IRF) of the detection channel. Several factors influence the IRF, in particular the transit time spread of the PMTs [17]. The goal of measuring the IRF is to obtain measurements that do not depend on the effects of the

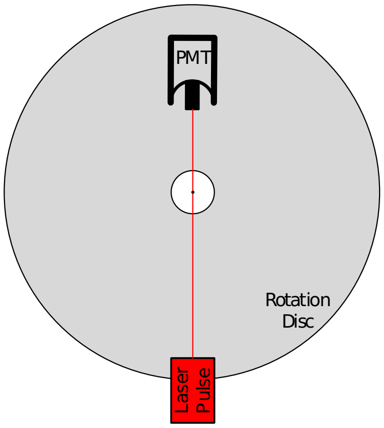


Figure 3.2 ***IRF measurement.***

measurement apparatus. In mathematical terms, a measured signal $y(t)$ obtained from a channel can be modeled as

$$y(t) = (h * x)(t) + n(t). \tag{3.1}$$

Here, $x(t)$ is the true signal from the object, $h(t)$ is the channel's IRF, and $n(t)$ is the noise on the measurement accounting for Poisson statistics associated with photon detection along with detector noise; $*$ is the temporal convolution operator.

Hence, to be able to measure the true signal, which is what a reconstruction algorithm uses, the IRFs of the scanner's channels must be measured in order to deconvolve them from the measured signals. To measure the IRF of a given channel, this channel is positioned in front of the laser source and IRFs are measured as depicted in Fig. 3.2. A neutral density filter is used to attenuate the laser pulse. The power of measured IRFs is not important here, since the aforementioned normalization is applied to TPSFs.

3.2.2 Relative calibration

In this step, all channels must be calibrated to an arbitrary detector (here, the first channel is considered as the reference channel). All measured TPSFs can be corrected for any observed differences by applying calibration factors for intensity and time delay to equalize and align all TPSFs. A homogeneous phantom is used as a diffuser and TPSFs are collected at certain angles. Then the channels are rotated as far as the angle difference between each adjacent channel (see Fig. 3.3). For example, channel 1 is rotated to be positioned at the angle of channel 2, and channel 2 is rotated to be positioned at the angle of channel 3 and so on. With this arrangement, channels 1 and 2, channels 2 and 3, . . . are calibrated to each other, and consequently, all channels can be calibrated to channel 1.

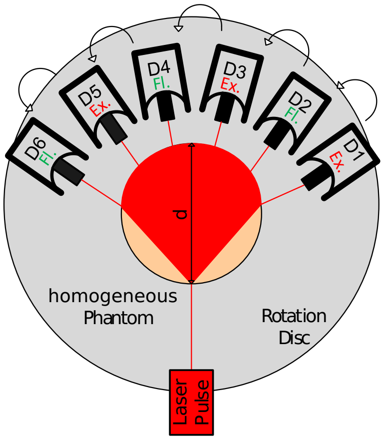


Figure 3.3 ***Relative calibration.***

To have more precise correction coefficients, the homogeneous phantom is rotated and translated to have more measurements. A total of 100 projections are acquired for this calibration. The final correction coefficients for the differences in detection, gain, and time delay are calculated as mean values of calculated coefficients at each projection.

3.3 Preprocessing

3.3.1 Dark noise subtraction for intrinsic channels

The level of dark noise is measured and subtracted from measured TPSFs. For this propose, the intrinsic channels (channels 1, 3, and 5 in Fig. 3.4) are isolated from any incoming light and a measurement is performed for a fixed collection time. Then the dark noise measurements are averaged over the collection time. The averaged dark level for each channel is subtracted from the measured TPSFs.

3.3.2 Leakage current subtraction for fluorescent channels

As for intrinsic channels, dark subtraction is carried out for the fluorescent channels. These dark signals may contain, another source of noise here because of potential excitation light leakage through the dichroic filters in the fluorescent channels (channels 2, 4, and 6 in Fig. 3.4). This combined leaked and dark signal must be subtracted from measured TPSFs. To measure the leak/dark signal for each channel, an optical phantom is used, and the signal in each channel is measured after diffusing through the optical phantom for a given collection time. The signals are then averaged over the collection time. The averaged noise signal for each channel is subtracted from the measured TPSFs. The measured leak/dark signals for intrinsic and fluorescent channels are depicted in Fig. 3.4.

3.3.3 TPSF Curve Smoothing

Since the image reconstruction algorithm uses each time bin of measured TPSFs, noise in the TPSFs can induce errors in the function/gradient evaluation. To suppress the noise, prior filtering of the TPSFs is carried out using a Savitzky-Golay filter, which is a popular filtering method in digital signal processing for smoothing out noisy data [133, 142]. The Savitzky-Golay filter uses a linear least squares function to fit a small window of data points to a low-degree polynomial. Then the polynomial estimates the point in the center of the window. Finally, the window moves forward by one data point and the process

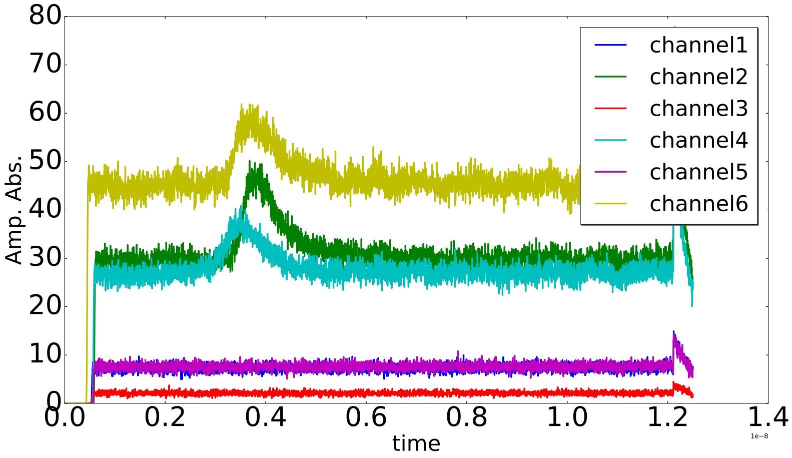


Figure 3.4 ***The measured dark and leakage currents.***

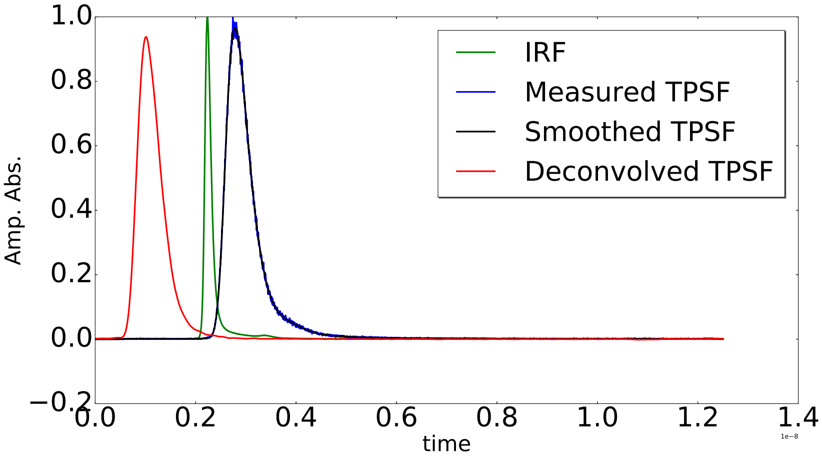


Figure 3.5 ***Sample curves in the Deconvolution process.***

is repeated. This continues until every point has been optimally adjusted relative to its neighbors.

3.3.4 Deconvolution

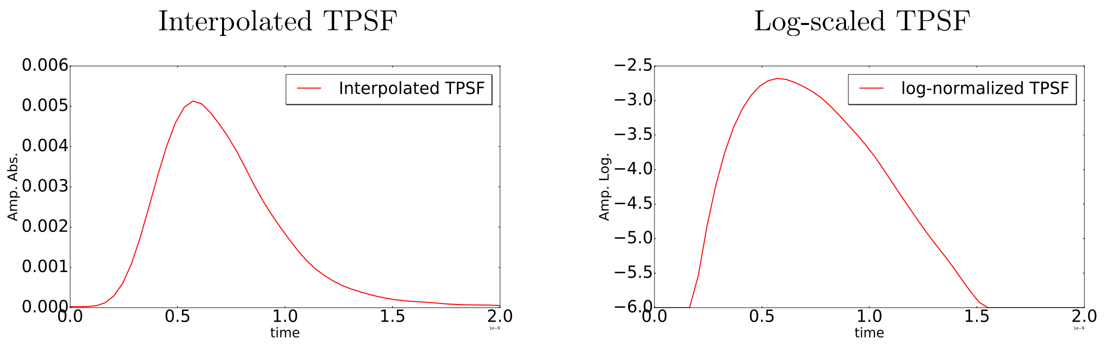
With the channels' IRFs at hand, it is possible to deconvolve the measured TPSFs from the measured IRFs (Fig. 3.5). For deconvolution, the Wiener deconvolution algorithm which performs well for noisy measurements is used [135]. This algorithm uses a Weiner filter in the frequency domain, to minimize the effect of noise at frequencies which have a low SNR. Based on Eq. (3.1), the goal is to calculate the $x(t)$. The Weiner algorithm allows obtaining an estimate $\hat{x}(t)$ of $x(t)$ by going to the frequency domain as follows:

$$\hat{X}(f) = Y(f)\frac{H^*(f)}{H(f)H^*(f) + \lambda^2}, \tag{3.2}$$

where λ is the SNR. In a Poisson distribution, if N is the number of counted photons in a defined time interval, then the standard deviation is $\sqrt{N}$ and SNR $= \dfrac{N}{\sqrt{N}} = \sqrt{N}$. Finally, by performing an inverse Fourier transformation on $\hat{X}(f)$, $\hat{x}(t)$ can be calculated.

3.3.5 TPSF normalization and scaling

After deconvolution, the deconvolved TPSFs are normalized with respect to their energies (integral of TPSFs). This normalization alleviates calibration difficulties as mentioned before. Also, down-sampling and scaling are performed on the deconvolved TPSFs before feeding them to reconstruction algorithm. The output of a TCSPC measurement carried out with the scanner is a curve with a length of 4096 time bins (time steps) that must be down-sampled to the number of time steps as defined in the forward problem. This number of time steps in the forward model is determined based on the optical properties


Interpolated TPSF

Log-scaled TPSF

Figure 3.6 ***Noise model and smooth filter outputs*** of the medium. For media considered in this work, a number between 50 to 75 is typical. Since the light intensity typically decays exponentially in media, a logarithmic scaling of the data is also applied, so as to keep the data in a range that remains in the same order of magnitude. Empirically, such scaling better conditions the optimization problem underlying the image reconstruction process; it is a standard technique in numerical optimization to scale data prior to optimizing a function [123]. In Fig. 3.6 the absolute value and log-scale values of a normalized TPSF and a down-sampled TPSF (50 time bins) is displayed.

This ends the description of the instrumentation that serve to acquire the data used in this thesis and the preprocessing thereof.

CHAPTER 4

MULTISPECTRAL IMAGE RECONSTRUCTION

4.1 Introduction

Diffuse optical tomography (DOT) is a relatively recent and non-invasive medical imaging technique in which near-infrared (NIR) light is utilized [80, 121]. DOT has the potential of being a low-cost imaging modality with acceptable accuracy with respect to other techniques [2]. The simplest approach to perform DOT imaging is through the use constant-amplitude light sources, called continuous-wave (CW)-DOT. CW-DOT has the advantages of low cost of installation and higher signal-to-noise ratio (SNR). The suitable wavelength range for optical imaging of biological tissues beyond depths of a few millimeters is between 650 to 930 nm [80], owing to the fact that absorption of the main chromophores (mainly oxy- and deoxy-hemoglobin) is least in that range. The major drawback of CW-DOT is its incapability of reconstructing simultaneously and distinguish the optical absorption and scattering coefficients, which are the main optical properties of biological tissues. The only solution to overcome the non-unicity of absorption and scattering recovery in CW-DOT is to resort to multispectral tomography [12]. In multispectral tomography, optical measurements are performed at several wavelengths. Herein, the dependency of tissue chromophore absorption and scattering factors to the wavelength is exploited to reconstruct the absorption and scattering properties simultaneously [32, 34]. The scale of the problem increases depending on the number of different chromophores considered along with the number of wavelengths used. This is the main difference of multispectral DOT compared to standard DOT as regards image reconstruction, as it might affect the optimization algorithm and consequently image reconstruction.

Image reconstruction in DOT is essentially a nonlinear problem due to the diffusion events in light propagation [165]. There are two distinct approaches to deal with this problem. In the first one, the problem is linearized using the first order Born or Rytov approximations and consequently, a linear system of equations can be solved iteratively [13, 127]. The second approach considers the nonlinear problem and attempts to solve it using the model fitting (parameter estimation). Herein, the model fitting problem in terms of the optical properties is solved through optimization. This is known as model-based iterative image reconstruction algorithm 2.3, which is discussed in general in 2.2.The solution is arrived at iteratively, and different optimization methods can be resorted to, including conjugate-gradient [85], quasi-Newton [96], Gauss-Newton [145], Levenberg-Marquardt [40], and truncated Newton [137] methods. All these methods use first order derivatives of the cost function, but depending on which one is resorted to, some will also exploit second order derivatives to minimize the cost function and find a solution [52].

In this study, we evaluate the performance of the aforementioned optimization methods in solving the multispectral DOT image reconstruction problem. To the best of the author's knowledge, this is the first time such a vast range of techniques is applied to the multispectral DOT problem. Considering the scale of the multispectral image reconstruction problem compared to that of the intrinsic DOT, one can expect from such a large-scale problem a more marked difference in the performance of the different optimization methods. The main intention in this work is to apply the most popular optimization techniques to multispectral DOT image reconstruction and to arrive at a recommendation as to which one should be favored. It should be noted that since all the aforementioned techniques have been implemented here, the standard algorithm for each of these techniques is considered along with some modifications to accelerate the execution of these techniques. The author is aware of the recent developments in optimization algorithms, but these sophisticated developments are out of the scope of this study. The mesh size, starting point (or initial guess), and noise power are three important factors that can affect the performance of an optimization algorithm. In this work, these different factors are considered for each optimization algorithm. Considering all these factors allows a comprehensive investigation of the behavior of each optimization algorithm for multispectral DOT image reconstruction.

This chapter is organized as follows: Section 4.2 presents the iterative model-based image reconstruction method used in this study. Section 4.3 details the optimization algorithms resorted to. Section 4.4 provides and displays the results for 3D numerical experiments. A comparison of reconstructed images is also presented in this section. Finally, section 4.5 discusses the optimization methods in multispectral reconstruction.

4.2 Image reconstruction

Iterative algorithms consist of two sub-problems. The first one called the *forward problem*, includes the modeling of light propagation inside the medium from the source to the detector locations. The result of the forward problem gives the radiant flux exiting the medium at its boundary based on its assumed optical properties. The second sub-problem, called the *inverse problem*, includes defining an appropriate cost function to measure the discrepancy between the experimental measurements and their predictions obtained from the forward problem, and minimizing this cost function iteratively through *optimization methods*. The predicted data from the forward problem is compared with the boundary measurements using the cost function at each iteration. Consequently, an update of the medium's optical properties is determined from the optimization method used [12, 35, 71, 90, 91, 119]. An iterative image reconstruction algorithm is terminated when the cost function reaches a predefined minimum criterion. The final representation of optical properties is then displayed as an image.

4.2.1 Forward problem

The most commonly used equation to describe the propagation of light in turbid media such as biological tissues is the diffusion equation (DE). In this work, the forward model is implemented in the CW domain using the DE given by

$$\mu_a(\mathbf{r})\Phi(\mathbf{r}) - \nabla \cdot [D(\mathbf{r})\nabla\Phi(\mathbf{r})] = Q(\mathbf{r}), \tag{4.1}$$

where $\mathbf{r}$ denotes position, $\mu_a(\mathbf{r})$ and $D(\mathbf{r})$ are the absorption and diffusion coefficients, $Q(\mathbf{r})$ is the source distribution and $\Phi(\mathbf{r})$ is the fluence rate. The diffusion coefficient is calculated based on the absorption coefficients μ_a, scattering coefficients μ_s and anisotropy factor (g) as $D(\mathbf{r}) = 1/[3(\mu_a(\mathbf{r})+(\mu_s(\mathbf{r})(1-g)))]$. The DE is accompanied with a boundary condition equation that the field (fluence rate) must satisfy on the boundary. Here the so-called refractive index mismatch boundary condition is used. It takes into account the change of refractive index that takes place at the boundary of a biological tissue and its surrounding medium (here air). This boundary condition takes the form

$$\left(\frac{1-R_\psi}{4}\right)\Phi\left(\mathbf{r}'\right) + \left(\frac{1+R_J}{2}\right)D(\mathbf{r}')\frac{\Phi\left(\mathbf{r}'\right)}{\hat{\mathbf{n}}} = b_T\left(\mathbf{r}'\right), \tag{4.2}$$

where R_ψ and R_J are reflection coefficients, $\hat{\mathbf{n}}$ is outward normal at the medium's boundary, and b_T is any source term located at the boundary [64].

Although not indicated above, the absorption coefficient has a wavelength dependency (absorption spectrum) given by [152]

$$\mu_a(\mathbf{r},\lambda) = \sum_{i=1}^{n} C_i(\mathbf{r})\varepsilon_i(\lambda), \tag{4.3}$$

where n is the number of chromophores (in the present case $n = 3$, since the main chromophores in biological tissues are oxy- and deoxy-hemoglobin (Hb, and HbO_2) and water (H_2O), $C_i(\mathbf{r})$ is the spatial concentration map of i-th chromophore and $\varepsilon_i(\lambda)$ is its extinction coefficient at wavelength λ. The $\mu_s' = \mu_s(\mathbf{r})(1-g)$ is called reduced scattering, and its wavelength dependency is well approximated by a power law as,

$$\mu_s'(\mathbf{r},\lambda) = A(\mathbf{r})\lambda^{-b((\mathbf{r})}, \tag{4.4}$$

in which A and b depend on the size, refractive indices and the concentration of scatterers [165]. In the NIR wavelength range, the so-called "scatter power" b can be considered constant as μ_s' does not vary much with wavelength in this range, and reconstruction can be performed for the spatially distributed scatter amplitude $A(\mathbf{r})$. In this work, b will thus be constant and equal to 1.3, which is a typical value for biological tissues [34].

To solve the forward problem, the DE must be discretized, leading to a set of algebraic equations that can be solved numerically. To obtain such equations, a grid or mesh must be defined over the medium and the variables of interest (fluence rate and the optical properties of the medium) are to be evaluated over this grid or mesh [139]. To this end, the finite difference method (FDM) is used here in the reconstruction process.

In DOT, the FDM is the simplest and fastest approach for finding the solutions of the forward problem [69, 93, 94]. Curved boundaries and complex geometries can be dealt with using the blocking-off method, which will be resorted to here [93]. The blocking-off approach has been shown to be accurate in modeling curved boundaries while preventing extended computational complexity [93].

4.2.2 Inverse problem

In this work, the discrepancy between boundary measurements and their predictions obtained from the forward model is quantified using the normalized least square cost function given by

$$f(\beta) = \frac{1}{2}\left\|\tfrac{\Phi^p}{\Phi^m} - 1\right\|_2^2 + \lambda\mathcal{R}(\beta), \tag{4.5}$$

where Φ^m and Φ^p are respectively the measured fluence rates and the fluence rates predicted by the forward model at all measurement nodes, β is the variables vector ($\beta = [C_1(r), C_2(r), C_3(r), A(r)]$, where C_i is the concentrations of the chromophore and A is the scatter amplitude) to be reconstructed, $\mathcal{R}$ is a penalty or regularization function of the unknowns, and λ is the regularization parameter. Several types of regularization terms have been considered in the literature, such as L^1 and L^2 norm regularization [9], incorporation of *a priori* information [46], or imposing smoothness on the spatial derivatives of the unknown vector [71]. Here, a first-order L^2 Tikhonov regularization [145] is used in the form,

$$\mathcal{R}(\beta) = \frac{1}{2}\left\|L(\beta_k - \beta_0)\right\|_2^2, \quad (L^TL)_{i,j} = \begin{cases} nn & \text{if } i = j, \\ -1 & \text{if } i, j \text{ in the defined neighborhood,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

Here nn is the number of voxel in the defined neighborhood. In this work, the 6 nearest neighbors define the neighborhood for each voxel in the 3D grid.

4.3 Optimization methods

In this study, as mentioned the multispectral DOT problem cast as an optimization problem written as

$$\min_{\beta\in R^n} \quad f(\beta) \quad \text{s.t.} \quad c_i(\beta) = 0, \quad i = 0, 1, 2, ..., m. \tag{4.7}$$

where $f : R^n \to R$ (objective function) and $c_i : R^n \to R^m$ (constraints) are twice continuously differentiable functions; n, m are the number of variables and constraints respectively. The necessary conditions for the point β^* to be considered as a minimum of the cost function f (at least locally) are that its gradient satisfies $g(\beta^*) = \nabla f(\beta^*) = 0$, and its Hessian H be positive definite, that is $H(\beta^*) = \nabla^2 f(\beta^*) \geq 0$. These conditions mean to assuming that f is strictly convex in the neighborhood of the optimal solution β^*. Numerical optimization algorithms can solve this convex problem iteratively to reach β^* by making available an approximation (current approximation) β_k at the k-th iteration and computing a new approximation β_{k+1} for the next iteration. Optimization techniques differ in the way a new approximation is obtained. The two main strategies to achieve this rely on line search and trust region techniques.

In the line search strategy, a new search direction (s_k) is obtained at each iteration, and an optimum point with smallest function value in this direction is searched for. In minimization problems, the search direction is a descent direction s_k, and the distance to move along can be found by approximately solving the following one-dimensional minimization problem:

$$\min \quad f(\beta_k + \alpha s_k), \quad \alpha > 0. \tag{4.8}$$

In this work the strong Wolfe condition [52] is used as criteria in the line search algorithm.

In the trust region strategy, a model function m_k (for example a second-order Taylor series expansion of the cost function) with similar behavior as the actual cost function f is constructed at the current point β_k [123]. The search at the current iteration is restricted to a minimizer of m_k within some region around β_k, called *trust region*, because the model m_k may not be a good approximation of f when β is far from β_k. In other words, the optimal step Δ is found by solving the following subproblem:

$$\min \quad m_k(\beta_k + \Delta), \quad \text{where } \beta_k + \Delta \text{ lies inside the trust region.} \tag{4.9}$$

The trust region at a given iteration k is chosen so that the model function m_k is a sufficiently accurate approximation of the cost function within that region, while allowing a sufficient decrease of the cost function. These two conditions generally play one against the other, because typically the smaller the region, the more accurate the approximation, whereas too small a region may lead to too small a decrease in the cost function. For this reason, the trust region typically needs to be adjusted within a given iteration [123].

The required quantities for most optimization techniques are the cost function f, the constraint functions c_i, and possibly the first derivatives (gradient g) and second derivatives (Hessian H) of f. Optimization techniques can be classified according to their use of the first and/or second order derivatives of the cost function. Important classes are nonlinear conjugate-gradient (CG) methods (use of first derivatives), Newton's methods (use of second derivatives) and quasi-Newton methods (use first derivatives and of an approximation of the Hessian). Good and popular references on optimization methods and their implementation are the book of Fletcher [52], and that of Nocedal and Wright [123]. Multispectral DOT image reconstruction is an under-determined nonlinear optimization problem. This means the problem has several possible solutions, and the convergence to a global minimum is not guaranteed. In this work, the two main optimization methods, that is Newton and quasi-Newton are considered. We split them into two families based on the implemented algorithm: i) gradient-based including Truncated Newton-CG (TNC) and quasi-Newton (L-BFGS) methods, and ii) Jacobian-based including Gauss-Newton (GN) and Levenberg-Marquardt (LM) methods. Standard CG methods, because of simplicity and slower convergence rate, are not considered in this work. In the following, the multispectral DOT inverse problem will be solved through 4 optimization algorithms and their performances will be compared.

4.3.1 Gradient-based family

Truncated Newton method

In Newton-CG method, to calculate the search direction and step length, the CG method and some of its variants are considered to handle negative curvature in the Hessian [123]. For more detail refer to [123]. In the TNC (also known as variants line-search Newton-CG method), the CG method is used to solve the Newton equations to find the search direction. The CG iteration is terminated as soon as a direction of negative curvature is generated [123]. This adaptation of the CG method produces a search direction s_k that is a descent direction. In fact, the CG algorithm is applied as the inner iteration to compute the search direction s_k at each iteration.

The CG will converge to the Newton step s_k if the approximated Hessian matrix B_k at a given iteration k is positive definite. The truncated Newton method does not require explicit full knowledge of the Hessian, but only matrix-vector product in the form $B_k \cdot v$ for any given vector v. This product can be calculated using the gradient vector via finite differencing. Therefore, the price for bypassing the computation of the Hessian is one new gradient evaluation per CG iteration.

Quasi-Newton methods

Table 4.1 ***Update formulas for gradient-based family***

	Update formula	Search direction calculation
TNC	$\beta_{k+1} = \beta_k - \alpha_k H_k^{-1} \cdot g_k$	solve $H_k \cdot s_k = -g_k$ by CG
L-BFGS	$\beta_{k+1} = \beta_k - \alpha_k B_k^{-1} \cdot g_k$	$B_k \cdot s_k = -g_k$ $B_{k+1} = B_k +$ $(y_k^T y_k)/(y_k s_k^T) - (B_k s_k s_k^T B_k)/(s_k^T B_k s_k)$ $y_k = g_{k+1} - g_k$

Quasi-Newton (QN) methods require only to calculate the first derivative of the cost function at each iteration. In comparison with Newton's method, since second derivatives are not required, quasi-Newton methods are sometimes more efficient [123]. Nowadays, the most popular quasi-Newton algorithms for solving unconstrained, constrained, and large-scale optimization problems are based on the low-memory BFGS method (L-BFGS). The update formula for BFGS methods is given in Table 4.1. An efficient way of implementing the update formula in the L-BFGS method is to use the inverse of the Hessian matrix at the first step and updating it at each iteration by applying the Sherman-Morrison formula [123] (the approximation of the Hessian H, is by tradition denoted by B).

4.3.2 Jacobian-based family

Gauss-Newton method

The GN method is the simplest method for minimizing the type of nonlinear least-square (LS) cost function as that used here for multispectral DOT. For this reason, most of the current reconstruction algorithms in DOT are implemented through GN methods. These methods can be viewed as a modification of Newton's method supplemented with a line search (also called damped Gauss-Newton). In the GN method, instead of solving Newton's equation $H_k \cdot s_k = -g_k$, an approximation of it is solved instead (Table 4.2). It saves from the costly direct calculation of the Hessian and calculation of the Jacobian matrix (J) can provide sufficient information about the descent direction [123]. After solving the modified Newton's equation, a line search in the descent direction leads to a reduced value of the cost function. The GN method update formula is shown in Table 4.2. In that formula, I represents the identity matrix and the μ is a scalar parameter that can be calculated by l-curve fitting or another strategy (l-curve is a heuristic technique to select the regularization parameter). The parameter μ brings the approximated Hessian to a positive definite region, which is helpful when the starting point is far from a minimum.

Levenberg-Marquardt method

The Levenberg-Marquardt (LM) method can be defined as an implementation of the Gauss-Newton (GN) method through the trust region technique [170]. The LM method can be obtained by using the same Hessian approximation in the GN method but replacing

Table 4.2 ***Update formulas for Jacobian-based family***

	Modified Newton's equation	Update formula
GN	$\left(J_k^T J_k + \mu I\right) \Delta\beta_k = -J_k^T r_k$	$\beta_{k+1} = \beta_k + \alpha_k \Delta\beta_k$
LM	$\left(J_k^T J_k + \mu_k G_k\right) \Delta\beta_k = -J_k^T r_k$	$\beta_{k+1} = \beta_k + \Delta\beta_k$

the line search with a trust-region strategy. The LM method can be useful when the Jacobian is inadequate or rank-deficient. Since the modified Newton equations are almost the same in both the GN and LM (Table 4.2), the local convergence properties of both methods can be expected to be similar. The LM method update formula is shown in Table 4.2. In that formula, G_k can be calculated with the equation $G_k = \big(\max(\mathrm{diag}(|J_k^T J_k|)) \times I\big)$ (I is identity matrix) and μ_k is a scalar parameter that can be calculated by l-curve fitting or another strategy and decreased monotonically at each step (here μ_k is calculated by l-curve fitting). Reducing the μ_k value (correspond to reduce the trust region area m_k) during the convergence to local minimum can improve the convergence process and make this method very robust in practice.

4.3.3 Algorithm implementation

All of the above optimization algorithms in two main categories (gradient-based and Jacobian-based) were implemented. In the gradient-based category only the value of the objective function f and the first derivatives (gradient) are required. In the Jacobian-based category, the calculation of the cost function and of the Jacobian matrix are required. The calculation of the gradient vector and of the Jacobian matrix are described in the following.

Gradient-based family implementation

In these algorithms, first, a forward model is solved to calculate the predicted measurements input to f. Then a convergence criterion to be satisfied by f will be checked. In the end, f should be small since it measures the discrepancy between the measurements and their predictions by the forward model, and thus such a criterion generally involves a threshold value denoted here ϵ_{thr}. If f does not meet this criterion, gradient is calculated through the adjoint state along with the Lagrangian method [167]. Consequently, the search direction is calculated according to the optimization method considered. Finally, a line search method is applied along the search direction to find the optimal step with the strong Wolfe conditions. This process is repeated until the criterion is satisfied (Algorithm 1). The algorithms in this category implemented in our reconstruction code are the L-BFGS and TNC. In this work, a bounded version of the L-BFGS algorithm [172] is used.

Algorithm 1 Gradient-based algorithm (L-BFGS, TNC)

```
procedure L-BFGS Function
    Define initial parameters ($\beta_0, \epsilon_{thr}$);
    Solve forward model with $\beta_0$;
    Compute $f_0, g_0$;
    $s_0 \leftarrow$ call L-BFGS;
    $\Delta f_k = f_0; k \leftarrow 0$;
    while $\left\| \frac{\Delta f_k}{f_k} \right\| \geq \epsilon_{thr}$ do
        $\alpha_k \Leftarrow$ Call line search function
        Compute $\beta_{k+1} = \beta_k + \alpha_k s_k$;
        Solve forward model for $\beta_{k+1}$;
        calculate $g_{k+1}$;
        $s_{k+1} \leftarrow$ call L-BFGS;                ▷ Table 4.1
        $\Delta f_k = f_k - f_{k-1}; k \leftarrow k + 1$;
    end while
```

```
procedure TNC Function
    Define initial parameters ($\beta_0, \epsilon_{thr}$);
    Solve forward model with $\beta_0$;
    Compute $f_0, g_0$;
    $s_0 \leftarrow$ call TNC;
    $\Delta f_k = f_0; k \leftarrow 0$;
    while $\left\| \frac{\Delta f_k}{f_k} \right\| \geq \epsilon_{thr}$ do
        $\alpha_k \Leftarrow$ Call line search function
        Compute $\beta_{k+1} = \beta_k + \alpha_k s_k$;
        Solve forward model for $\beta_{k+1}$;
        calculate $g_{k+1}$;
        $s_{k+1} \leftarrow$ call TNC;                   ▷ Table 4.1
        $\Delta f_k = f_k - f_{k-1}; k \leftarrow k + 1$;
    end while
```

Jacobian-based family implementation

In these algorithms, first, forward model is solved to calculate the predicted measurements and the cost function value. Then a convergence criterion is evaluated and, if the criterion is not satisfied, the gradient and the Hessian approximations are calculated through the calculation of the Jacobian matrix from reciprocal theory [10]. The LM algorithm in this work is implemented through a trust-region scheme. This means the hyperparameter μ_k from Table 4.2 changes at each iteration. It is reduced successively if the function decrease condition ($f_{k+1} < f_k$) is satisfied, otherwise it is increased until the reduction is satisfied. The modified Newton equation for the LM algorithm is solved with iterative solvers like LGMRES which is an accelerated version of the GMRES solver that uses the Krylov subspace method (GMRES stands for generalized minimal residual, and LGMRES for "loose" GMRES [15]). For the GN method, the modified Newton equation is implemented in a Jacobian-Free-Newton-Krylov (JFNK) subspace and solved with the LGMRES solver [98]. To further accelerate the LGMRES solver for solving the Newton equation in both algorithms, an approximation of the Hessian matrix inversion is provided through the L-BFGS formula. Both algorithms are shown in Algorithm 2. In these algorithms, the first derivative is calculated and then the LGMRES solver is called and the modified Newton equation is solved for each algorithm in Table 4.2. In this study, the values 10^{-4} and 10^{-3} are used for μ and μ_0 in the GN and LM respectively. For a starting point close to the local minimum, a smaller value (here 10^{-4}) is used and for a starting point far from the local minimum, a larger value (here 10^{-3}) is used.

Algorithm 2 Jacobian-based algorithm (LM, GN)

```
procedure LM Function
    Define initial parameters $(\beta_0, \epsilon_{thr}, \mu_0)$;
    Solve forward model with $\beta_0$;
    Compute $f_0, J_0, G$;
    $\Delta\beta_0 \leftarrow$ call LM;                                ▷ Table 4.2
    $\Delta f_k = f_0; k \leftarrow 0$;
    while $\left\| \frac{\Delta f_k}{f_k} \right\| \geq \epsilon_{thr}$ do
        Compute $\beta_{k+1} = \beta_k + \Delta\beta_k$;
        Solve forward model for $\beta_{k+1}$;
        if $f_{k+1} < f_k$ then
            calculate $J_{k+1}$;
            $\mu_{k+1} \leftarrow \mu_k / \mu_{down}$;
        else
            $\beta_{k+1} \leftarrow \beta_k$;
            $J_{k+1} \leftarrow J_k$;
            $\mu_{k+1} \leftarrow \mu_k * \mu_{up}$;
        $\Delta\beta_{k+1} \leftarrow$ call LM;                        ▷ Table 4.2
        $\Delta f_k = f_k - f_{k-1}; k \leftarrow k + 1$;
    end while
```

```
procedure GN Function
    Define initial parameters $(\beta_0, \epsilon_{thr}, \mu)$;
    Solve forward model with $\beta_0$;
    Compute $f_0, J_0$;
    $\Delta\beta_0 \leftarrow$ call GN;                                ▷ Table 4.2
    $\Delta f_k = f_0; k \leftarrow 0$;
    while $\left\| \frac{\Delta f_k}{f_k} \right\| \geq \epsilon_{thr}$ do      ▷ Criterion for function value
        $\alpha_k \Leftarrow$ Call line search function
        Compute $\beta_{k+1} = \beta_k + \alpha_k \Delta\beta_k$;
        Solve forward model for $\beta_{k+1}$;
        calculate $f_{k+1}, J_{k+1}$;
        $\Delta\beta_{k+1} \leftarrow$ call GN;                        ▷ Table 4.2
        $\Delta f_k = f_k - f_{k-1}; k \leftarrow k + 1$;
    end while
```

It is worth mentioning here that when an optimization problem with several parameters is faced, such as is the case here with multispectral DOT reconstruction, a parameter scaling is necessary especially for the L-BFGS algorithm (Eq. (4.10)). This parameter scaling is helpful in making the convergence rate faster. One alternative for parameter scaling could be a gradient/Jacobian scaling ([34]) but, in this study, parameter scaling showed better performance in comparison with gradient/Jacobian scaling. For Newton's method, parameter scaling could be neglected as it does not require scaling, but nevertheless, scaling is applied for all implemented algorithms. Explicitly, the parameter scaling used here is as follows:

$$\beta = \left[\frac{C_1(r)}{\max(c_1(r))}, \frac{C_2(r)}{\max(c_2(r))}, \frac{C_3(r)}{\max(c_3(r))}, \frac{A(r)}{\max(A(r))} \right]. \tag{4.10}$$

4.4 Results

In this section, the results of the multispectral DOT image reconstruction algorithm using gradient-based and Jacobian-based optimization methods are presented. A numerical cylinder phantom is used for this study. The performance of optimization algorithms for this numerical phantom is evaluated for different conditions. The scale of the problem, the starting point (initial guess), and the noise level are the 3 important factors that are considered for both gradient-based and Jacobian-based optimization methods. The results are evaluated in terms of iteration steps and computation time. In addition, the image quality of the results is evaluated with two image metrics, namely the correlation and deviation factors.

The correlation factor $\rho_a \in [-1, 1]$ of the reconstructed image is determined with respect to the target image. This factor is calculated as [95]:

$$\rho_a = \frac{\sum_{i=1}^{i=N} \left(\beta_i^r - \hat{\beta}_i^r\right) \left(\beta_i^t - \hat{\beta}_i^t\right)}{(N-1)\sigma_{\beta^r}\sigma_{\beta^t}}, \tag{4.11}$$

where

$$\sigma_{\beta^t} = \sqrt{\frac{1}{N-1}\sum_{1}^{N}(\beta_i^t - \hat{\beta}_i^t)^2}, \quad \sigma_{\beta^r} = \sqrt{\frac{1}{N-1}\sum_{1}^{N}(\beta_i^r - \hat{\beta}_i^r)^2}. \tag{4.12}$$

Here β^r and β^t are the reconstructed and target images, N is the number of voxels, σ_{β^t} and σ_{β^r} are the standard deviations of the target and reconstructed images, and finally $\hat{\beta}_i^r$ and $\hat{\beta}_i^t$ indicate the mean values of the reconstructed and target images respectively. Values of ρ_a close to 1 show a high correlation between the reconstructed and target images (high accuracy). For small values of ρ_a, low accuracy and no correlation between the target and reconstructed images is obtained.

The deviation factor $\rho_b \in [0, \infty]$ to evaluate the image quality is defined as follows:

$$\rho_b = \frac{(1/N)\sum_{i=1}^{N}\left(\beta_i^r - \beta_i^t\right)^2}{\sigma_{\beta^t}}. \tag{4.13}$$

As its name implies, this factor determines a metric to measure the deviation of the reconstructed image from the target image. A small value of ρ_b indicates a reconstructed image with high accuracy.

All of the reconstruction results were obtained on a desktop computer with an Intel quad Core i7-2700K CPU running at 3.50 GHz × 4 and with 32 GB RAM. Programming was done in Python 2.7 (Numpy/Scipy [81, 159] packages) and Cython.

4.4.1 Numerical experiment

Several 3D numerical experiments are conducted on a numerical cylindrical phantom with a 25 mm diameter and 40 mm height to simulate the size of a small mouse. The experiments are conducted for 2 types of chromophores - Hb and HbO_2, water (H_2O), and scatterer (4 different optical parameters in total). For each of optical parameter considered here, an inclusion with different sizes is positioned at different locations (4 inclusion in total). The

3D view of the phantom and each concentration map are shown in Fig. 4.1. The following set of wavelengths is selected for this study: 650 nm, 718 nm, 880 nm, 930 nm. These wavelengths and associated extinction coefficients (ϵ_i) for each inclusion are presented in Table 4.3. This set of wavelengths is chosen so that the greatest changes are observed in the values of the extinction coefficients. A 10 μM concentration is used as the background concentration for Hb and HbO_2 and 20 μM is used for the concentration of the Hb and HbO_2 inclusions; 18% and 36% are used as the background and inclusion concentrations for water, and 4.5 (10^{-3}mm^{-1}) and 9 (10^{-3}mm^{-1}) as the background and inclusion scatter amplitude.

Table 4.3 ***Molar extinction coefficients of chromophores (ϵ_i) and absorption coefficient of water used in the experiment.***

	650 nm	718 nm	880 nm	930 nm
$\epsilon_{[HbO_2]}$ (mm.M)$^{-1}$	368	340	1170	1222
$\epsilon_{[Hb]}$ (mm.M)$^{-1}$	3750.12	1368.28	736.6	763.84
$\mu_{a[H_2O]}$ (mm)$^{-1}$	0.0003594	0.0009913	0.0055978	0.017296

The numerical experiments are designed to investigate each optimization algorithm under several conditions. Source and detector arrangements are configured at a plurality of positions around the phantom. The measurements are acquired around the phantom at different heights between 10 mm and 30 mm at steps of 2 mm along the phantom (10 slices in total). Eight source locations (projections) are used to illuminate the phantom at each height with an angle of 45° between each source location. Sixteen detection locations collect the light from 100° to 260° with respect to the illumination direction (transillumination regime) at steps of 10°. The total number of projections is $4 \times 8 \times 10 = 320$ which leads to collecting a total of $4 \times 8 \times 16 \times 10 = 5120$ measurements.

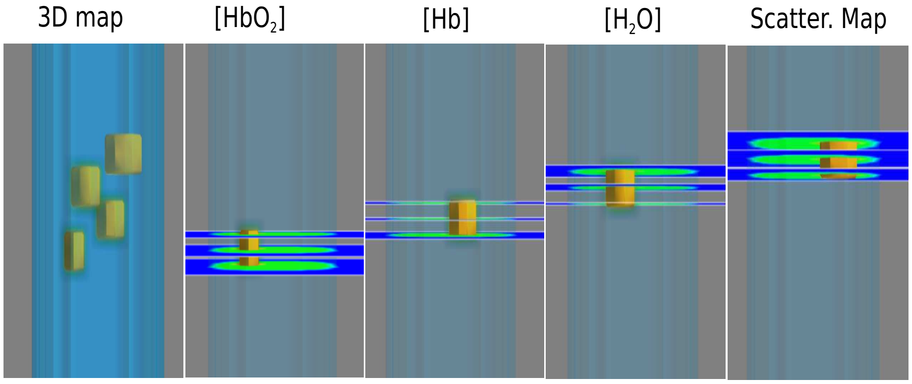


Figure 4.1 ***3D numerical phantom with one inclusion for each optical parameter (Hb, HbO_2, H_2O, and scatterer). Four different optical parameters in total are positioned at different heigth.***

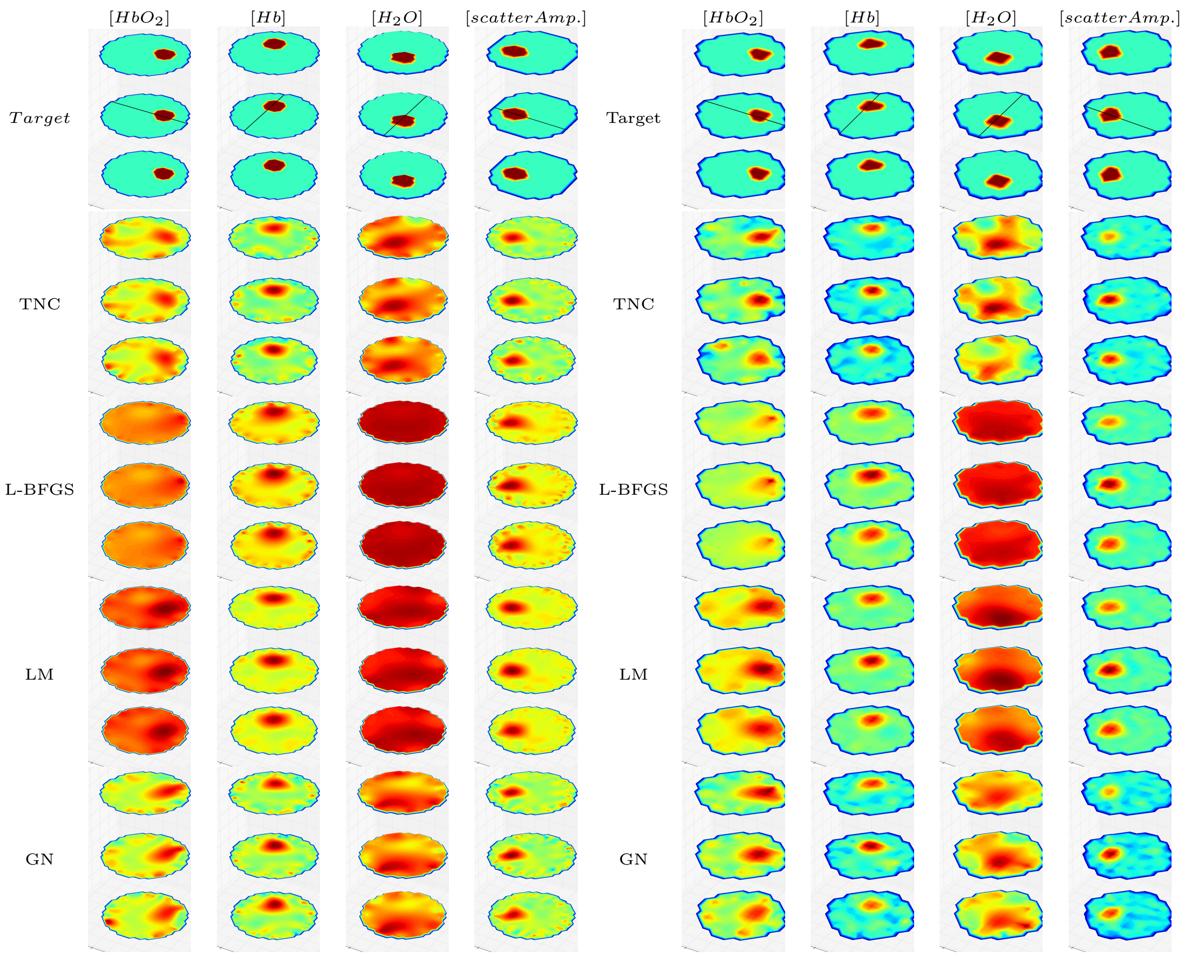
The numerical experiments allow investigating optimization methods in multispectral DOT image reconstruction in terms of the performance and the quality of reconstructed images. Six numerical experiments are designed to evaluate 3 important factors, namely: scale of the problem, starting point, and noise level. Two numerical simulations for different mesh sizes (fine and coarse mesh), 2 numerical simulations for different starting points (20% above and below of background value), and 2 numerical simulations for different noise levels (20 dB and 10 dB signal-to-noise ratios (SNR)) are considered. For all the optimization methods the same criterion is used as shown in Algorithm 2 and Algorithm 1. The threshold parameter is selected to be $\epsilon_{thr} = 10^{-3}$ for this work. The results obtained are evaluated for each condition in sequence in the following.

4.4.2 Scale of the problem

An important factor in an optimization problem is its scale. In multispectral DOT image reconstruction, a fine spatial discretization increases the scale of the inverse problem. To investigate this factor, two mesh sizes are considered, a fine and a coarse. For the fine mesh, the numerical phantom is discretized into a structured-grid with voxel size 1 mm × 1 mm × 1 mm and a total of 19,240 nodes. For the coarse mesh, the voxel size is 2 mm × 2 mm × 2 mm for a total of 2,420 nodes. The fine mesh size results in 4 × 19,240 $\sim$ 80,000 variables and 4 × 2,420 $\sim$ 10,000 variables. One can see that the scale of the problem for the fine mesh is much larger than that for the coarse mesh. For the fine mesh, almost 80,000 variables must be reconstructed from 5120 measurements, while for the coarse mesh 10,000 variables must be reconstructed from 5120 measurements, that is indicative of the difference in problem ill-posedness in these two cases. To test for the other conditions, the phantom is also discretized into a mid-scale to have a trade-off between the ill-posedness and spatial resolution. The voxel size in this case is 1.5 mm × 1.5 mm × 1.5 mm for a total of 5018 nodes. This discretization makes the parameter variable β a vector of size 4 × 5018 $\sim$ 20,000. All of the synthetic data were generated on a structured-grid with voxel size 0.75 mm × 0.75 mm × 0.75 mm to avoid the inverse crime.

Table 4.4 ***Results for fine mesh.***

	$f_0 = 57.73$			$[HbO_2]$		$[Hb]$		$[H_2O]$		Scatterer	
	f_{fin}	n_{iter}	$time(s)$	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b
TNC	0.142	14	8865	0.31	1.0	0.51	0.9	0.24	1.01	0.51	0.84
L-BFGS	0.170	25	3756	0.28	0.97	0.51	0.86	0.23	0.99	0.55	0.84
LM	0.169	15	2406	0.39	0.92	0.61	0.80	0.22	0.98	0.69	0.79
GN	0.142	7	4994	0.36	0.97	0.49	0.91	0.24	0.99	0.48	0.89



(a) Fine mesh, 20 dB. **(b)** Coarse mesh, 20 dB.

Figure 4.2 ***Reconstructed concentration maps for 2 different scales: fine mesh and coarse mesh.***

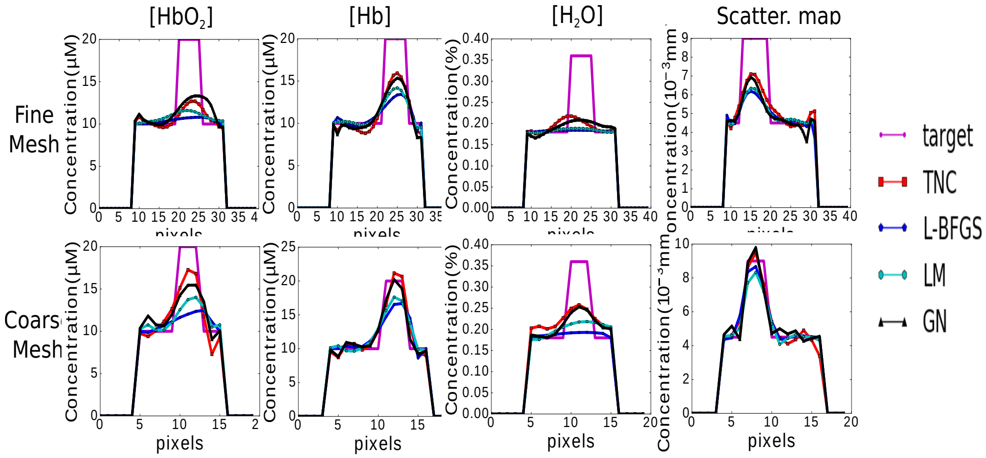


Figure 4.3 ***Contrast line of a 2D slice for 2 different scales: fine mesh and coarse mesh.***

Table 4.5 ***Results for coarse mesh.***

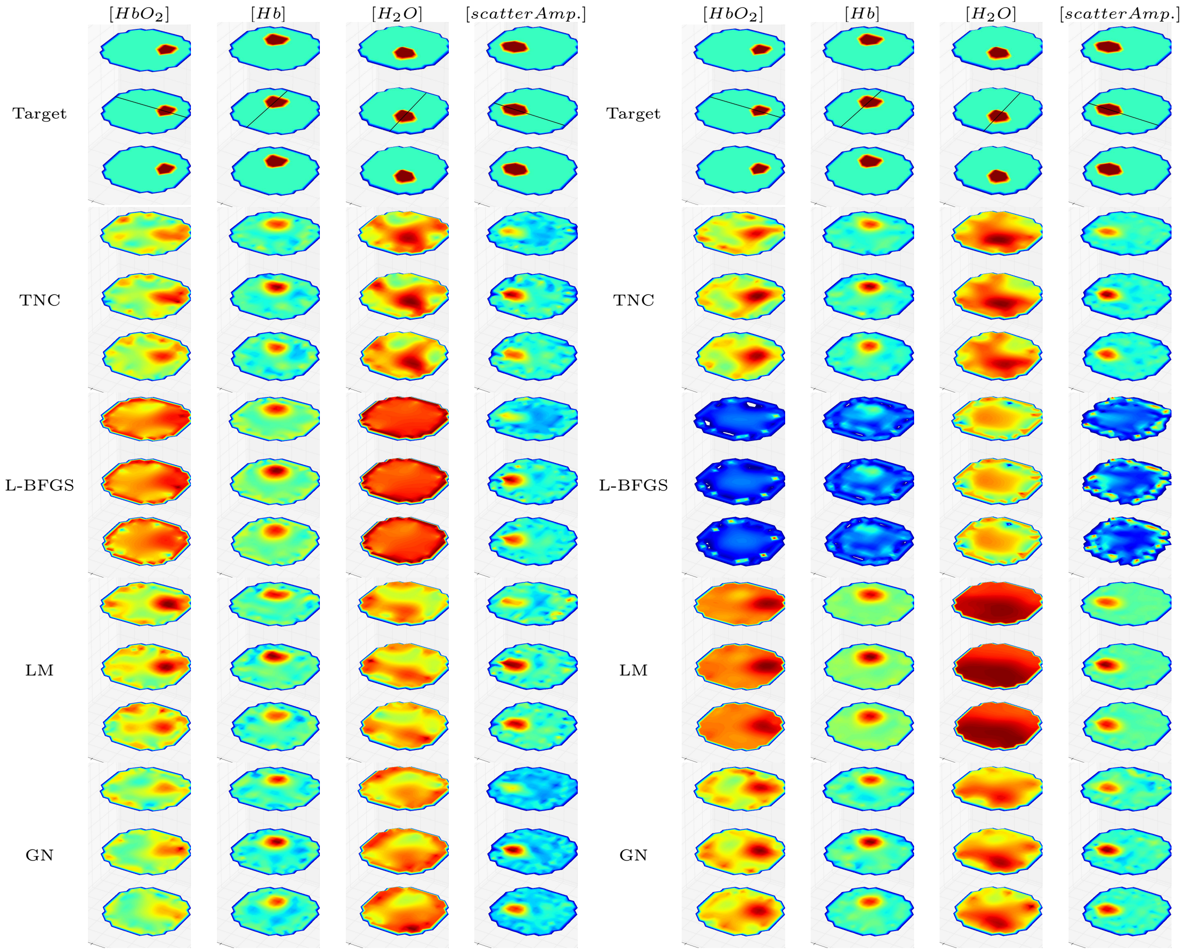
	$f_0 = 55.67$			$[HbO_2]$		$[Hb]$		$[H_2O]$		Scatterer	
	f_{fin}	n_{iter}	$time(s)$	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b
TNC	0.169	12	163	0.53	0.94	0.70	0.75	0.35	1.06	0.70	0.72
L-BFGS	0.197	25	77	0.43	0.90	0.70	0.71	0.34	0.96	00.80	0.60
LM	0.187	7	165	0.55	0.83	0.76	0.65	0.37	0.92	0.82	0.59
GN	0.179	6	166	0.44	0.98	0.65	0.83	0.41	0.95	0.65	0.82

Fine mesh Reconstructed concentration maps obtained from the TNC and GN optimization methods are better as shown in Fig. 4.2a. In fact, these methods converge to a lower minimum with the final value of the objective function (f_{fin}) being less that for the other two methods, as can be seen in Table 4.4. Also, the correlation factor (ρ_a) and the deviation factor (ρ_b) are larger and smaller respectively for the GN and TNC. Overall, the results are not good as for the coarse mesh as expected because of the ill-posedness of the problem.

Coarse mesh All results for the coarse mesh are shown in Fig. 4.2b and Table 4.5. The TNC and GN terminate at a smaller objective function value in comparison with the two other methods. In comparing the fine mesh with the coarse mesh, the reconstructed inclusion values for the coarse mesh are closer to the target values as the line contrast shows in Fig. 4.3. All concentration maps are well located except water for the L-BFGS and LM. In fact, these methods perform better when they approach to a local minimum. The reason is that water has the smallest changes (smallest gradient norm) among the other variables with respect to the selected wavelengths. Therefore, for water, good results as for the other variables are not expected.

4.4.3 Starting point

The starting point determines the convergence path to the local/global minimum. Convergence may significantly vary from one starting point to another. In this work, the optimization algorithms are evaluated for two starting points that are significantly distant from one another to better simulate real experiments. It was decided to define the initial map 20% above or 20% below the target background. These two choices are far from each other and could happen in real experiments because of possible errors in providing the initial guess.



(a) +20%, 20 dB. **(b)** -20%, 20 dB.

Figure 4.4 ***Reconstructed concentration maps for 2 different starting points: 20% above and under the background.***

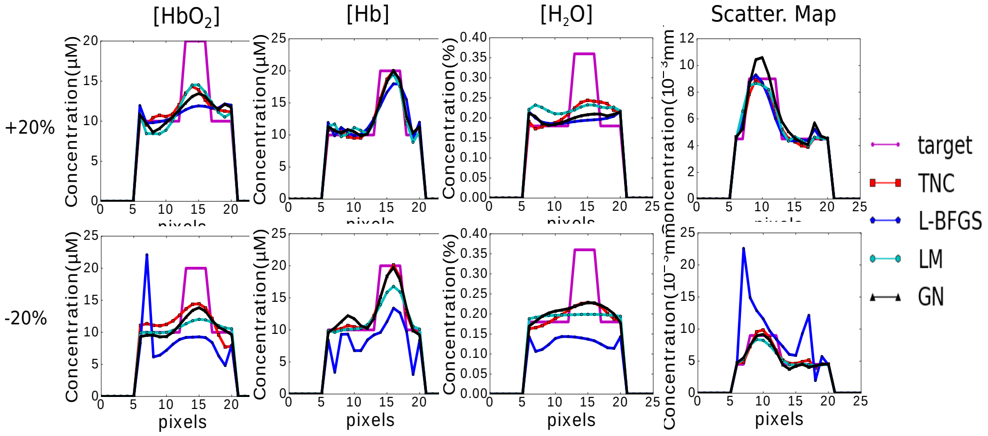


Figure 4.5 ***Contrast line of a 2D slice for 2 different starting points: 20% above and under the background.***

Table 4.6 ***Results for +20%.***

	$f_0 = 25.92$			$[HbO_2]$		$[Hb]$		$[H_2O]$		Scatterer	
	f_{fin}	n_{iter}	$time(s)$	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b
TNC	0.172	13	920	0.26	1.25	0.59	0.96	0.30	1.12	0.54	0.97
L-BFGS	0.188	40	498	0.12	1.26	0.53	1.01	0.05	1.24	0.53	0.94
LM	0.237	7	775	0.36	1.07	0.58	0.94	0.25	1.08	0.58	0.91
GN	0.167	7	471	0.24	1.19	0.58	0.93	0.11	1.15	0.63	0.86

Table 4.7 ***Results for -20%.***

	$f_0 = 239.23$			$[HbO_2]$		$[Hb]$		$[H_2O]$		Scatterer	
	f_{fin}	n_{iter}	$time(s)$	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b
TNC	0.188	18	990	0.42	1.00	0.71	0.74	0.36	0.96	0.73	0.70
L-BFGS	161.7	40	412	0.06	3.43	0.23	2.66	0.10	2.57	0.02	9.27
LM	0.200	17	658	0.41	0.91	0.75	0.67	0.26	0.96	0.83	0.59
GN	0.163	12	946	0.41	0.96	0.68	0.75	0.32	0.98	0.71	0.71

+20% The results for 20% above the real background value (+20%) are shown in Fig. 4.4a and Table 4.6. By looking at the reconstructed images and the calculated image metrics, the results of all methods are very similar, with a difference in water and HbO_2 concentration. These two have smaller gradient vectors in comparison to the scatter amplitude and Hb for the selected wavelengths. For water and HbO_2, the LM and TNC show better results and larger correlation factor (ρ_a) values.

-20% The initial function value (f_0) for 20% below the real background value (-20%) is almost 10 times larger than the +20% in this work. However, the results for all optimization methods are close to the results for +20% except for L-BFGS. For this starting point, the L-BFGS method fails to converge properly to a reasonable minimum. Others starting points were considered, for example (-15%), and then the L-BFGS algorithms converge properly (the results are not displayed here), but for values larger than 15% the L-BFGS fails to converge. One reason for this is that quasi-Newton methods can be trapped in a saddle point in comparison with Newton's methods. For example, a saddle point could exist close to this starting point that the L-BFGS cannot pass and, therefore, falls into this saddle point. The results for -20% for all methods are shown in Fig. 4.4b and Table 4.7.

4.4.4 Noise level

Measurements in DOT are primarily contaminated with noises that are usually modeled as Poisson or Gaussian distributions. In the case of CW light imaging, normally the power and the number of detected photons is greater than for other methods, and the noise can be modeled as a Gaussian distribution. In this work, to model noise in the measurements, additive white Gaussian noise (AWGN) is selected. AWGNs with powers of 1% and 10% of the measurements are added directly to the synthetic measurement data (equal to 20 dB and 10 dB SNR respectively).

20 dB SNR All methods show similar results for the concentration of Hb and the scatterer amplitude and a difference can be seen for water and the concentration of HbO_2, as one can see better results obtained with the LM algorithm (Fig. 4.6a). One reason for these observations is that the LM method is implemented with a trust-region method. This pushes the approximated Hessian to a positive definite regime to find a descent direction. Moreover, when the minimum is approached, a smaller trust region is used and the algorithm can converge deeper as can be seen from the value of f_{fin}, which is the smallest among the values of the other methods (Table 4.8).

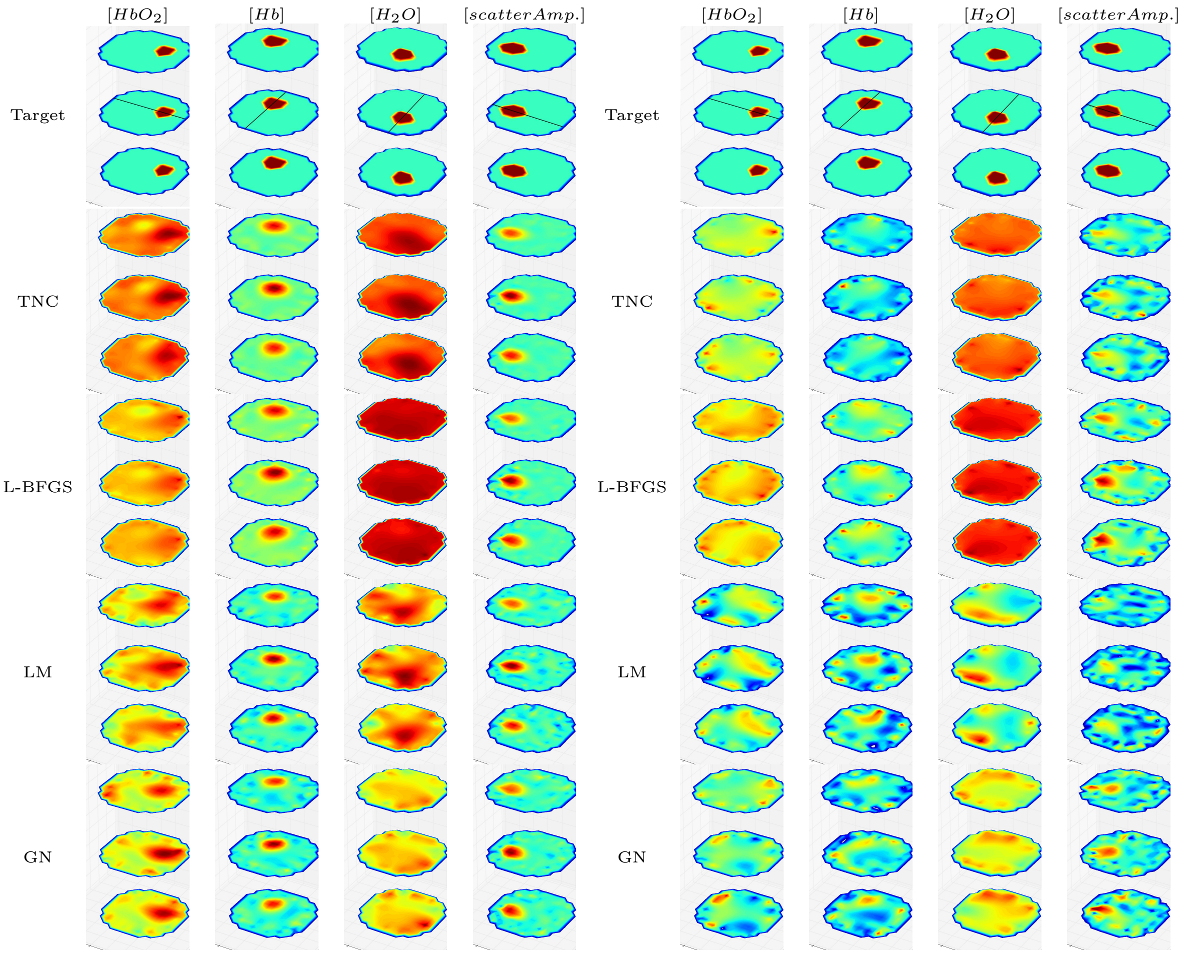
10 dB SNR For 10 dB SNR, the worst SNR in practical experiments, the LM and L-BFGS give better results. Since the LM is implemented with a trust-region approach, it shows robustness to noise and finds a descent direction (Fig. 4.6b). In the L-BFGS, the step is smaller than Newton's methods and is smoother, because the approximate Hessian is calculated from a few last iterations of the gradient vector. This means the accumulation of noise in the Hessian matrix, and consequently in finding the descent direction is smaller than in the case of Newton's methods. Also, one can see smaller fluctuations in the reconstructed values in the L-BFGS results in comparison with other algorithms.

4.5 Conclusion

In this study, quasi-Newton optimization methods (L-BFGS) and Newton's optimization methods (TNC, LM, GN) were evaluated for multispectral DOT. The optimization methods were sorted based on their implementation requirements into 2 main categories: gradient-based and Jacobian-based. In gradient-based families including L-BFGS and TNC, the gradient of the objective function is calculated with respect to the variables

Table 4.8 ***Results for 20 dB SNR.***

	$f_0 = 24.73$			$[HbO_2]$		$[Hb]$		$[H_2O]$		Scatterer	
	f_{fin}	n_{iter}	$time(s)$	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b
TNC	0.195	13	460	0.39	0.92	0.77	0.64	0.36	0.93	0.85	0.56
L-BFGS	0.191	35	310	0.32	0.94	0.71	0.70	0.30	0.97	0.79	0.61
LM	0.169	9	857	0.33	1.05	0.68	0.76	0.35	0.99	0.68	0.75
GN	0.173	8	499	0.42	0.96	0.67	0.79	0.15	1.05	0.67	0.76



(a) 20 dB SNR. **(b)** 10 dB SNR.

Figure 4.6 ***Reconstructed concentration maps for 2 different noise level: 20 dB and 10 dB.***

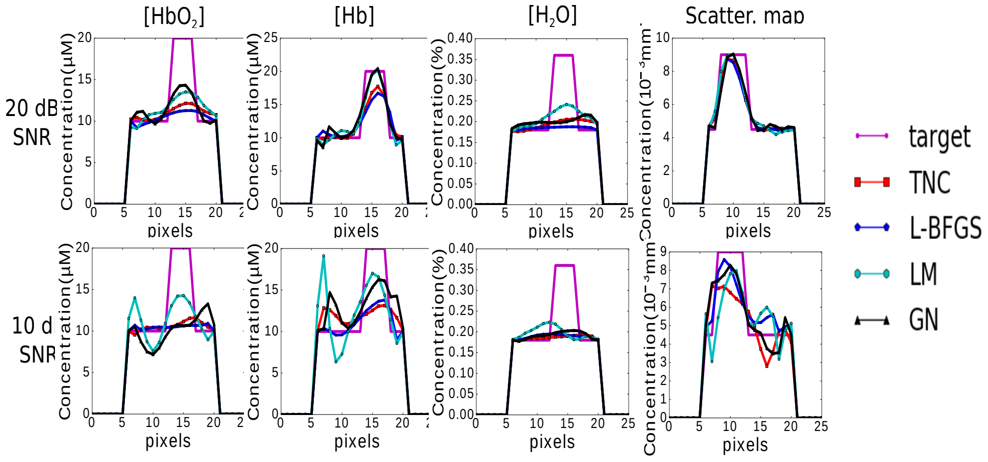


Figure 4.7 ***Contrast line of a 2D slice for 2 different noise level: 20 dB and 10 dB.***

Table 4.9 **Results for 10 dB SNR.**

	$f_0 = 88.37$			$[HbO_2]$		$[Hb]$		$[H_2O]$		Scatterer	
	f_{fin}	n_{iter}	$time(s)$	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b	ρ_a	ρ_b
TNC	19.32	5	382	0.12	1.11	0.21	1.29	0.12	0.99	0.27	1.44
L-BFGS	19.65	9	372	0.10	1.04	0.29	1.13	0.25	0.97	0.35	1.36
LM	18.94	4	717	0.19	1.71	0.29	1.75	0.02	1.46	0.24	2.04
GN	19.20	5	430	0.02	1.74	0.23	1.79	0.00	1.41	0.29	1.81

by the adjoint technique. The optimal steps in all algorithms in this category, are obtained through a line search. In the Jacobian-based family including LM and GN, the Jacobian matrix is calculated with respect to the variables to be reconstructed by applying reciprocal theory. For the Jacobian-based family, the modified Newton's equation is solved via a Jacobian-Free Newton-Krylov (JFNK) approach that is very efficient for a large problem such as multispectral DOT reconstruction problem. In the Jacobian-based family, the adequate selection of the hyperparameter λ is important. A large value of λ makes the convergence slow while a small one leads to failure in finding the optimal step. The precise choice of this hyperparameter is critical and can be calculated with l-curve methods. A parameter scaling was considered for all algorithms. The computational time of the gradient vector is N_d times (N_d = number of detectors) faster than the computational time of the Jacobian matrix while the step in the Jacobian-based family is quadratic and the number of function evaluation is smaller than for the gradient-based family. The most time consuming step in the Jacobian-based family is solving the Newton's equation which is implemented with a JFNK technique to accelerate it. The image quality of reconstructed images were evaluated by observing the reconstructed images and using two image metrics, correlation and deviation factors. Furthermore, the performance of the algorithm is evaluated in terms of final cost function values, the total number of basic function evaluations, and the total computational time.

The optimization algorithms were investigated for 3 important factors, the scale of the problem, starting point, and noise level in several conditions. In practical experiments, one always deals with different SNR measurements because of different acquisition strategies and equipment non-ideality. Also, providing the initial guess is always a difficult task and errors in the calculation of the initial guess is admissible. Investigation of optimization methods under several conditions can help choosing the right methods to enhance the image quality of reconstructed images.

In multispectral DOT image reconstruction, several concentrations can be recovered simultaneously. The selection of a series of wavelengths that provide large differences for all concentrations is almost impossible. Therefore, some of the concentrations have larger gradient norms in comparison with other concentrations. In this work, the scatter amplitude, Hb, HbO_2, and water have the largest gradient norm respectively (Table 4.3). Since the contribution of chromophores and scatterer in the gradient vector is different because of changes in molar extinction to the light wavelength, a different reconstructed image for each chromophore is expected. All of the optimization methods converge to an acceptable reconstructed image for two concentrations with larger gradient norms (here scatterer and Hb). For the two concentrations with the smallest gradient norms (here HbO_2 and water), some differences are observed between the performance of the optimization algorithms. For both fine and coarse meshes, TNC and GN lead to better reconstructed concentrations in comparison with other algorithms. For different initial maps, TNC and LM reconstructed better result for $+20\%$, and for -20% TNC and GN recovered better results. The L-BFGS algorithm for -20% fails to converge to an acceptable local minimum. On reason for this behaivor can be referred to the fact that quasi-Newton methods can be trapped in a saddle point. In noisy conditions, the LM and L-BFGS reconstruct better concentration map. These differences in the performance of optimization algorithms tell us to use optimization algorithms based on the specifics of the problem. However, the decision to use which optimization algorithm rests with the user and should be dependent on the problem to be solved.

Generally, optimization algorithms with line search technique perform well for nonlinear problems such as multispectral DOT image reconstruction. Under very noisy conditions, the descent direction contaminates with noise and reaching a descent direction is difficult. In these conditions, the algorithms using the trust-region method such as LM are more robust. Also, trust-region methods work well when the initial guess is close to the minimum. Since, the behavior of optimization algorithm will be changed with respect to the initial map therefore, in a real multispectral experiment, if we do not have precise prior information it is recommended to repeat the reconstruction process for several initial maps to obtain accurate images.

CHAPTER 5

FLUORESCENCE LIFETIME IMAGE RECONSTRUCTION

5.1 Preface

Title:
Combined continuous-wave/time-domain image reconstruction for fluorescence lifetime tomography

Authors and affiliation:

- SeyedRohollah Ettehadi: PhD student, Université de Sherbrooke, Faculté de génie, Département de génie électrique et de génie informatique, Laboratoire TomOptUS.
- Y. Bérubé-Lauzière : Professeur (directeur de doctorat), Université de Sherbrooke, Faculté de génie, Département de génie électrique et de génie informatique, Laboratoire TomOptUS.

Submission date:
8 December 2017

Journal:
Journal of Computerized Medical Imaging and Graphics

Contribution of the document to the thesis

This article proposes a novel image reconstruction algorithm for fluorescence lifetime tomography. The proposed image reconstruction algorithm extracted all information of the measured temporal curve in the reconstruction process, which enable us to reconstruct several fluorescence lifetimes simultaneously. Full-curved reconstruction algorithms has not been developed because of imposing a computationally demanding reconstruction time. The proposed algorithm accelerates this process by a two steps image reconstruction.

Abstract:

Time-domain (TD) fluorescence lifetime tomography reconstruction using full-curve time point-spread functions (TPSF) is computationally very expensive, owing to the need of time stepping over each TPSF. In this work, a combined continuous-wave (CW)/TD image reconstruction algorithm for fluorescence lifetime tomography is presented. It allows to accelerate TD data-based image reconstruction and make it more practical. In the proposed algorithm, the map reconstructed from CW data is thresholded and used as a fluorescence yield mask for the tomographic reconstruction of fluorescence lifetime using the TD data. The algorithm is validated with numerical experiments. The results demonstrate the ability of the algorithm to provide maps distinguishing fluorophores having different lifetimes. In the case of a complex mouse phantom, the algorithm converges to an accurate image in a reasonable computation time.

5.2 Introduction

Fluorescence diffuse optical tomography (FDOT), as a medical imaging technique for functional and molecular imaging of biological tissues, has seen important developments in the past decade [39]. Fluorescence imaging relies on an external source of light, typically a laser, to excite a fluorescent compound, followed by the detection of the ensuing emitted light signal. Fluorescence as a contrast mechanism in optical imaging enhances the capability of differencing between normal and diseased tissues [65].

The two main quantities of interest in FDOT are the fluorescence yield and the fluorescence lifetime (FLT) [108]. The yield relates to the fluorophore's concentration, and thus is indicative of where it accumulates. This is of key importance for imaging fluorescent probes designed to target specific diseases in tissues, *e.g.* cancer. As regards the FLT, it presents several interesting features: (1) Fluorescent probes with overlapping emission spectra but different FLTs can be discriminated and used concurrently [99, 100]. (2) FLT varies with local environment biochemical and physical factors (pH, O_2 level, presence of Ca^{2+}, temperature), thus allowing to sense such factors [19, 67]. This is widely used in microscopy [153] and is the basis for FLT imaging (FLIM), which also allows robust measurements of molecular interactions at the nanometer scale within living cells and tissues through FRET (Förster resonance energy transfer) [77]. FRET is exploited to develop smart activatable probes [166] which light up when cleaved by a specific enzyme. *In vivo* FRET tomography in small animals is a promising modality for monitoring the more efficient release of drugs to targeted diseased tissues and reduce secondary damage [161], and for studying fundamental disease-related enzymatic reactions [113]. 3D *in vivo* FLT imaging has the potential to become a key tool in pharmacology and biology.

Most FDOT systems are based on continuous-wave (CW) measurements that require less expensive instrumentation in comparison with frequency-domain (FD) and time-domain (TD) measurements. However, CW measurements only allow recovering the fluorescence yield, while FD and TD systems are capable of providing both the fluorescence yield and the FLT. TD systems measure the temporal point spread function (TPSF) at the boundary of a medium. This offers the richest information among the three measurement techniques, since a laser pulse virtually contains all modulation frequencies, including the CW (zero frequency) measurement.

Several image reconstruction algorithms for FLT tomography have been developed in recent years. Most of them use FD measurements or a Fourier/Laplace transform of TPSFs obtained from TD measurements, and reconstruct the FLT from FD (complex) data [9, 54, 59, 124, 128, 129]. Other algorithms use the TPSF curves and reconstruct the FLT from TD data [7, 29, 73, 106, 109, 120]. A comparison of these two approaches has been carried out by Kumar [105], who showed that TD approaches can separate multiple lifetime targets significantly better than FD approaches. FLT tomography based on full temporal curves requires stacking TPSFs in a measurement vector. Such a method can be computationally very demanding, as it requires to step in time over the whole of each TPSF, and this for all TPSFs. Further, this needs to be repeated at each iteration of the image reconstruction algorithm. Instead of using full curve information, some algorithms use some aspects of TPSFs such as the slope of the TPSFs' tail [106], early/late time bins of TPSFs [73], or the moments of TPSF [7, 29, 109]. However, using early time bins or time gates does not exploit the full information contained in TPSFs, and some information is lost for the sake of computational efficiency. These methods could lead to a limitation and inability to perform high-throughput imaging of multiple fluorescent targets in FLT tomography. The maximum number of discriminated distinct fluorophores that are reported in these algorithms is three. While representing a higher computational challenge, using full curve information of measured TPSFs can lead to more robust algorithms for unmixing and localization of more fluorophores.

In this work, a nonlinear image reconstruction algorithm is developed for reconstructing the FLT spatial distribution of fluorophores through a combined CW/TD approach to reduce the computational load. First, a CW reconstruction algorithm provides a map of the distribution of the fluorescence yield. This map is then used in a second stage as a mask for full-curve TD image reconstruction in regions of interest as determined by the yield mask. This second stage resorts to a gradient-based method for recovering the FLT of fluorophores. A nonlinear algorithm is considered here as the FLT image reconstruction problem is intrinsically nonlinear.

This paper is organized as follows. Section 2 discusses model-based image reconstruction and associated requirements pertaining to the present work. Section 3 provides the details of the algorithm developed and implemented. Numerical experiments along with results validating the proposed algorithm are described in Section 4. Finally, Section 5 concludes the paper.

5.3 Model-Based Image Reconstruction

FLT tomography based on full temporal curves requires stacking TPSFs in a measurement vector. The algorithm developed here follows the model-based image reconstruction

(MBIR) scheme [6, 90]. First, a light propagation model predicts the light distribution in the medium under examination. Then, a least-square data fitting function, so-called the objective function (OF), is used to assess the discrepancy between experimental measurements and predictions thereof at a set of detector locations. The OF is minimized through an iterative optimization algorithm.

5.3.1 Forward Model

The forward problem in FDOT predicts the light flux at detector locations on the surface of a medium from a given distribution of: absorption and scattering coefficients, the fluorescence yield, and the FLT throughout the medium. The general model for light propagation inside biological tissues is the radiative transfer equation (RTE). However, owing to the complexity of solving the RTE, the diffusion equation (DE) is ubiquitous in biomedical optics. The DE is a partial differential equation (PDE) derived from the RTE using the diffusion approximation [164], which assumes that the light field propagating in the medium is isotropic at each point. The DE has proved to be a very useful model and is adopted here. For practical geometries with complex curved boundaries as those encountered in biomedical optics (*e.g.* the shape of a mouse in preclinical molecular imaging), the DE cannot be solved analytically, and numerical methods are resorted to. Modeling the propagation of excitation and fluorescence light in a biological tissue requires two coupled equations, which when using the DE as a model are as follows for a medium Ω and its boundary $\partial\Omega$ [24]:

$$\begin{aligned}\frac{1}{\nu}\frac{\partial\Phi_x}{\partial t} + \left(\mu_a^x(\mathbf{r}) + \mu_a^{x\to m}(\mathbf{r})\right)\Phi_x(\mathbf{r},t) - \nabla\cdot\left[D^x(\mathbf{r})\nabla\Phi_x(\mathbf{r},t)\right] &= Q_x(\mathbf{r},t), \quad \mathbf{r}\in\Omega,\\ \frac{1-R_\psi}{4}\Phi_x(\mathbf{r},t) + \frac{1+R_J}{2}D^x(\mathbf{r})\frac{\partial\Phi_x}{\partial\hat{\mathbf{n}}} &= 0, \quad \mathbf{r}\in\partial\Omega,\end{aligned} \tag{5.1a}$$

$$\begin{aligned}\frac{1}{\nu}\frac{\partial\Phi_m}{\partial t} + \mu_a^m(\mathbf{r})\Phi_m(\mathbf{r},t) - \nabla\cdot\left[D^m(\mathbf{r})\nabla\Phi_m(\mathbf{r},t)\right] &= Q_m(\mathbf{r},t), \quad \mathbf{r}\in\Omega,\\ \frac{1-R_\psi}{4}\Phi_m(\mathbf{r},t) + \frac{1+R_J}{2}D^m(\mathbf{r})\frac{\partial\Phi_m}{\partial\hat{\mathbf{n}}} &= 0, \quad \mathbf{r}\in\partial\Omega.\end{aligned} \tag{5.1b}$$

Here x and m are indices referring to the excitation and emission wavelengths; Φ_x and Φ_m are the light fluences (W/cm^2), and μ_a^x and μ_a^m are the absorption coefficients at the excitation and emission wavelengths, $\mu_a^{x\to m}$ is the absorption coefficient of excitation light by the fluorophore, D^x and D^m are the diffusion coefficients, Q_x and Q_m are the excitation and fluorescence sources, and $\mathbf{r}$ is the coordinate vector of a point. The index of refraction is assumed to be the same at both wavelengths. Hence, the speed of light in the medium is the same at both wavelengths, and is denoted by ν. The quantities R_ψ and R_J are coefficients related to the refractive index mismatch at the boundary of the medium, and are related to the Fresnel reflection coefficient [64]. The unit normal vector at the boundary is denoted by $\hat{\mathbf{n}}$. Typically, in time-domain measurements, the excitation source $Q_x(t)$ is an ultra-short laser pulse which can be modeled as a Dirac delta function. The fluorescence source $Q_m(\mathbf{r}, t)$ is obtained from the excitation fluence, fluorescence yield and FLT of the fluorophore as

$$Q_m(\mathbf{r}, t) = \frac{\eta}{\tau(\mathbf{r})} \mu_a^{x \to m}(\mathbf{r}) \int_0^t \Phi_x(\mathbf{r}, t') \exp\left(-\frac{t - t'}{\tau(\mathbf{r})}\right) dt', \tag{5.2}$$

where $\eta\mu_a^{x\to m}$ and τ are respectively the fluorophore's yield and FLT, both of which may vary spatially. To solve the forward problem, a finite difference scheme is used to spatially discretize the light propagation equations. For a time-domain problem as considered here, the time derivatives in Eq. (5.1) are also discretized by a finite difference approach, leading to the following iterative algebraic linear equations

$$\begin{aligned}(\theta A_{x,m}(\boldsymbol{\beta}) + B)\, \boldsymbol{\Phi}_{x,m}(t_n) = {} & [(\theta - 1) A_{x,m}(\boldsymbol{\beta}) + B]\, \boldsymbol{\Phi}_{x,m}(t_{n-1}) \\ & + \mathbf{Q}_{x,m}(t_{n-1}), \quad t_n > t_0,\end{aligned} \tag{5.3}$$

with the initial condition

$$\boldsymbol{\Phi}_{x,m}(t_0) = 0, \quad t_0 = 0. \tag{5.4}$$

Here, $\boldsymbol{\Phi}_{x,m}$ are the vectors of the spatially discretized excitation and emission fluences and similarly for $\mathbf{Q}_{x,m}$, the time steps are specified as $t_{n+1} = t_n + \Delta t$, $A_{x,m}$ are matrices of coefficients pertaining to the excitation and emission equations resulting from the finite difference discretization, B is a diagonal matrix with value $1/(\nu\Delta t)$ on its diagonal. The vector $\boldsymbol{\beta}$ represents the optical parameters of the fluorescence problem, *i.e.* fluorescence yield and FLT (the intrinsic (background) absorption μ_a^x, μ_a^m, and scattering D^x, and D^m are considered to be known; see [53] and the Results section below. The parameter θ controls the temporal discretization scheme; it can be set to a value in the range from 0 (explicit scheme) to 1 (full implicit scheme). In this work, $\theta = 1/2$ is used, corresponding to the stable Crank-Nicholson scheme customarily used with parabolic PDEs as those considered here. Solving these equations gives the evolution in time of the excitation ($\boldsymbol{\Phi}_x$) and emission ($\boldsymbol{\Phi}_m$) light fluences spatial distributions.

5.3.2 TPSF Normalization and Scaling

The TPSFs are normalized with respect to their energy before being feed to the OF. This normalization has a great impact in practice, as the amplitude of a measured TPSF is affected by several factors, such as the quantum efficiency of the detectors, the optical design of the collection channels, the out-of-focus conditions for the collection lens with respect to the medium's surface, *etc.* It is technically very difficult to calibrate optical instruments so that their values they output are absolute. To avoid such a difficulty, which is significant in practice, it is advantageous to be able to work with TPSFs whose amplitudes are normalized in some way. TPSFs normalized by their energy, can be considered as probability distribution functions (PDFs), and this is how they will be normalized here. Thus, in fact the OF will fit the predicted PDFs to the measured PDFs. Since the light intensity typically decays exponentially in media, a logarithmic scaling of the data is also applied after normalization, so as to keep the data in a range that remains in the same order of magnitude. Empirically, such scaling better conditions the optimization problem underlying the image reconstruction process; it is a standard technique in numerical optimization to scale data prior to optimizing a function [123]. The final predicted and measured TPSFs after normalization and scaling are defined as

$$\begin{aligned} \hat{\boldsymbol{\Phi}}_m(t) &= \ln\left(\frac{\Phi_m(t)}{E_{\Phi_m}}\right), \quad E_{\boldsymbol{\Phi}_m} = \int_0^T \boldsymbol{\Phi}_m(t)dt, \\ \hat{M}(t) &= \ln\left(\frac{M(t)}{E_M}\right), \quad E_M = \int_0^T M(t)dt, \end{aligned} \tag{5.5}$$

where, $\boldsymbol{\Phi}_m$ is a TPSF predicted by the light propagation model, E_{Φ_m} and E_M are the energies of their respective TPSFs, and $\hat{\boldsymbol{\Phi}}_m$ and $\hat{M}$ are the predicted and measured TPSFs after normalization and scaling. It is worth to mention that the measured TPSFs are deconvolved prior to being used in Eq. (5.5).

5.3.3 Inverse Model

A nonlinear inverse problem such as that considered here, which is an ill-posed and underdetermined problem, is generally solved using iterative optimization techniques [28]. To do this, a least square data fitting function (objective function - OF), is used to model the discrepancy between measured TPSFs and predicted TPSFs. The OF used here is defined as

$$f\left(\boldsymbol{\beta}\right)=\sum_{i=1}^{N_s}\sum_{k=1}^{N_t}\frac{1}{2}\left\|\frac{P_i\hat{\boldsymbol{\Phi}}_{m,i}(\boldsymbol{\beta},t_k)-\hat{M}_i(t_k)}{\sigma_i(t_k)}\right\|_2^2+\lambda\mathcal{R}(\boldsymbol{\beta}), \tag{5.6}$$

constrained by

$$\begin{cases} F_x(\boldsymbol{\beta})\boldsymbol{\Phi}_{x,i}(t_k)=\boldsymbol{Q}_{x,i}(t_k), \\ F_m\boldsymbol{\Phi}_{m,i}(t_k)=\boldsymbol{Q}_{m,i}(\boldsymbol{\beta},t_k). \end{cases} \tag{5.7}$$

Here i is an index to denote source locations, P_i is a projection operator to obtain the predicted measurements at the detector locations for each tomographic projection i (*i.e.* the set of measurements made when the source is at the location with index i), $\hat{\boldsymbol{\Phi}}_{m,i}$ and $\hat{M}_i$ are respectively the predicted normalized fluoresced fluence field and the normalized measured set of TPSFs for source i, $\sigma_i(t_k)$ is the standard deviation of the i-th measurement at t_k (here because of the Poisson nature of measurements, it is equal to $\frac{1}{2}\hat{M}_i(t_k)$), $\boldsymbol{\beta}$ is the vector of parameters to be reconstructed (fluorescence yield and FLT at each voxel), λ is the regularization hyperparameter, $\mathcal{R}$ is the regularization function, and N_s, N_t are the number of projections (source locations) and time bins respectively. The OF defined here, which is constrained as indicated by the forward model equations adjoining the OF and appearing in Eq. (5.7), depends on the vector of parameters $\boldsymbol{\beta}$. The forward model equations have been written in operator form (F_x and F_m) in Eq. (5.7) to simplify notations, but in the implementation of these equations in the computer program, they are used in the form of Eq. (5.3). This OF indirectly depends on $\boldsymbol{\beta}$ (which is reminded in Eq. (5.5) by including $\boldsymbol{\beta}$ as a parameter in $\hat{\boldsymbol{\Phi}}_{m,i}$), a fact that makes the inverse problem a nonlinear one. To minimize the OF, Newton or gradient-based optimization strategies can be resorted to. Newton methods are not suitable for time-domain measurements, because calculation and storage of the sensitivity matrix (Jacobian) and of the Hessian matrix is too much computationally and memory-wise intensive. For these reasons a gradient-based optimization method is called for here to minimize the OF [123].

5.3.4 Gradient-Based Optimization

Gradient-based optimization methods, only require the computation of the function value (f) and its first derivatives, that is the gradient denoted by ($\frac{\partial f}{\partial \beta}$). Here, the low-memory BFGS (L-BFGS) quasi-Newton optimization algorithm resorted to. This is similar to the line search Newton method, the main difference being that an approximation of the inverse Hessian is used in place of the exact one. Additionally, an efficient representation of the approximate inverse Hessian further reduces significantly the computational burden. For details on the L-BFGS algorithm, the reader is referred to [123].

5.3.5 Gradient Calculation

Gradient-based optimization algorithms, as their name implies, require the calculation of the gradient of the OF. An efficient way to calculate the gradient is to use an adjoint method [28], whereby a Lagrange function is resorted to. In this context, the Lagrange function consists of the OF augmented by the light propagation equations as constraints via Lagrange multipliers (also called adjoint variables). In the present case, the Lagrange takes the form

$$\begin{aligned} L\left(\boldsymbol{\beta}\right) = f\left(\boldsymbol{\beta}\right) &+ \sum_{i,k=1}^{N_s,N_t} \langle \left(F_x(\boldsymbol{\beta})\boldsymbol{\Phi}_{x,i}(t_k) - \boldsymbol{Q}_{x,i}(t_k)\right), \boldsymbol{\Psi}_{x,i}(t_k)\rangle \\ &+ \sum_{i,k=1}^{N_s,N_t} \langle \left(F_m\boldsymbol{\Phi}_{m,i}(t_k) - \boldsymbol{Q}_{m,i}(\boldsymbol{\beta}, t_k)\right), \boldsymbol{\Psi}_{m,i}(t_k)\rangle. \end{aligned} \tag{5.8}$$

Here the $\boldsymbol{\Psi}_x$'s and $\boldsymbol{\Psi}_m$'s are the adjoint variables associated with the excitation and emission PDEs respectively. The benefit of such an approach is that the total work of computing the OF and its gradient ($\partial f/\partial\boldsymbol{\beta}$) is approximately equivalent to solving one time step of the forward model [25].

From the theory of Lagrange multipliers, the Lagrange function must be stationary, leading to the following equation:

$$\begin{aligned} \delta L = \langle\frac{\partial f}{\partial\boldsymbol{\beta}}, \delta\boldsymbol{\beta}\rangle + \sum_{i,k=1}^{N_s,N_t}\langle\frac{\partial L}{\partial\boldsymbol{\Psi}_{x,i}(t_k)}, \delta\boldsymbol{\Psi}_{x,i}(t_k)\rangle + \sum_{i,k=1}^{N_s,N_t}\langle\frac{\partial L}{\partial\boldsymbol{\Psi}_{m,i}(t_k)}, \delta\boldsymbol{\Psi}_{m,i}(t_k)\rangle + \\ \sum_{i,k=1}^{N_s,N_t}\langle\frac{\partial L}{\partial\boldsymbol{\Phi}_{x,i}(t_k)}, \delta\boldsymbol{\Phi}_{x,i}(t_k)\rangle + \sum_{i,k=1}^{N_s,N_t}\langle\frac{\partial L}{\partial\boldsymbol{\Phi}_{m,i}(t_k)}, \delta\boldsymbol{\Phi}_{m,i}(t_k)\rangle = 0 \end{aligned} \tag{5.9}$$

Since the variations $\delta\boldsymbol{\Psi}_{x,i}(t_k)$, $\delta\boldsymbol{\Psi}_{m,i}(t_k)$, $\delta\boldsymbol{\Phi}_{x,i}(t_k)$, $\delta\boldsymbol{\Phi}_{m,i}(t_k)$, and $\delta\boldsymbol{\beta}$ can be considered independent (the Lagrange multipliers can be chosen to achieve this), the stationarity equation leads to the following set of equations:

$$\begin{cases} \frac{\partial L}{\partial\boldsymbol{\Psi}_{x,i}(t_k)} = F_x(\boldsymbol{\beta})\boldsymbol{\Phi}_{x,i}(t_k) - \boldsymbol{Q}_{x,i}(t_k) = 0 & \textit{excitation state eq.} \\ \frac{\partial L}{\partial\boldsymbol{\Psi}_{m,i}(t_k)} = F_m\boldsymbol{\Phi}_{m,i}(t_k) - \boldsymbol{Q}_{m,i}(\boldsymbol{\beta}, t_k) = 0 & \textit{emission state eq.} \\ \frac{\partial L}{\partial\boldsymbol{\Phi}_{x,i}(t_k)} = \frac{\partial f}{\partial\boldsymbol{\Phi}_{x,i}(t_k)} - F_x(\boldsymbol{\beta})^T\boldsymbol{\Psi}_{x,i}(t_k) = 0 & \textit{excitation adjoint eq.} \\ \frac{\partial L}{\partial\boldsymbol{\Phi}_{m,i}(t_k)} = \frac{\partial f}{\partial\boldsymbol{\Phi}_{m,i}(t_k)} - F_m^T\boldsymbol{\Psi}_{m,i}(t_k) = 0 & \textit{emission adjoint eq.} \\ \frac{\partial L}{\partial\boldsymbol{\beta}} = \frac{\partial f}{\partial\boldsymbol{\beta}} + \sum_{i,k=1}^{N_s,N_t}\langle\frac{\partial(F_x(\boldsymbol{\beta})\boldsymbol{\Phi}_{x,i}(t_k))}{\partial\boldsymbol{\beta}}, \boldsymbol{\Psi}_{x,i}(t_k)\rangle - \\ \sum_{i,k=1}^{N_s,N_t}\langle\frac{\partial\boldsymbol{Q}_{m,i}(\boldsymbol{\beta},t_k)}{\partial\boldsymbol{\beta}}, \boldsymbol{\Psi}_{m,i}(t_k)\rangle + \lambda\frac{\partial\mathcal{R}}{\partial\boldsymbol{\beta}} = 0 & \textit{optimality condition.} \end{cases} \tag{5.10}$$

We observed that the contribution of the adjoint parameters related to the excitation equation (the $\boldsymbol{\Psi}_x$'s) to the computation of the gradient $\partial f/\partial\boldsymbol{\beta}$ via the optimality condition is much smaller than that of the $\boldsymbol{\Psi}_m$'s. Hence, the adjoint equation for excitation and its contribution to the calculation of the gradient can be neglected, thus further reducing the computational burden. Obtaining adjoint parameters from an adjoint equation is computationally similar to solving a state equation except for the fact that the adjoint parameters are calculated backward in time, *i.e.* from the latest time to $t = 0$. Finally, the system of Eqs. (5.10) can be rewritten as the following set of equations:

$$\begin{aligned} &\boldsymbol{\Phi}_{x,i}(t) = F_x(\boldsymbol{\beta})^{-1}\boldsymbol{Q}_{x,i}(t), \quad \boldsymbol{\Psi}_{m,i}(t) = (F_m)^{-1}\frac{\partial f}{\partial \boldsymbol{\Phi}_{m,i}(t)}, \\ &\boldsymbol{\Phi}_{m,i}(t) = F_m^{-1}\boldsymbol{Q}_{m,i}(\boldsymbol{\beta},t), \quad \frac{\partial f}{\partial\boldsymbol{\beta}} = -\sum_{i,k=1}^{N_s,N_t}\langle\frac{\partial \boldsymbol{Q}_{m,i}(\boldsymbol{\beta},t_k)}{\partial\boldsymbol{\beta}}, \boldsymbol{\Psi}_{m,i}(t_k)\rangle - \lambda\frac{\partial\mathcal{R}}{\partial\boldsymbol{\beta}}, \end{aligned} \tag{5.11}$$

where the term $\frac{\partial Q_{m,i}(\boldsymbol{\beta},t)}{\partial\boldsymbol{\beta}}$ and $\frac{\partial f}{\partial\Phi_{m,i}}$ can be obtained through Eqs. (5.2), (5.5) and (5.6) as follows:

$$\frac{\partial \boldsymbol{Q}_{m,i}(\boldsymbol{\beta},t)}{\partial\boldsymbol{\beta}_j} = \begin{cases} \frac{1}{\tau(\mathbf{r}_j)}\int_0^t \boldsymbol{\Phi}_{x,i}(\mathbf{r}_j,t')\exp\Big(-\frac{t-t'}{\tau(\mathbf{r}_j)}\Big)dt', \ \boldsymbol{\beta}_j = \eta\mu_a^{x\to m}(\mathbf{r}_j) \\ \frac{\eta\mu_a^{x\to m}(\mathbf{r}_j)}{\tau^2(\mathbf{r}_j)}\int_0^t\Big(-1+\frac{t-t'}{\tau(\mathbf{r}_j)}\Big)\boldsymbol{\Phi}_{x,i}(\mathbf{r}_j,t')\exp\Big(-\frac{t-t'}{\tau(\mathbf{r}_j)}\Big)dt', \boldsymbol{\beta}_j = \tau(\mathbf{r}_j), \end{cases} \tag{5.12}$$

where $\mathbf{r}_j \in \Omega$, and

$$\frac{\partial f}{\partial\boldsymbol{\Phi}_{m,i}} = \frac{\partial f}{\partial\hat{\boldsymbol{\Phi}}_{m,i}}\frac{\partial\hat{\boldsymbol{\Phi}}_{m,i}}{\partial\boldsymbol{\Phi}_{m,i}} = P_i^T(\frac{P\hat{\boldsymbol{\Phi}}_{m,i}(\boldsymbol{\beta},t) - \hat{M}_i(t)}{\sigma^2(t)}) \circ \frac{E_{\boldsymbol{\Phi}_{m,i}} - \boldsymbol{\Phi}_{m,i}}{E_{\boldsymbol{\Phi}_{m,i}} \circ \boldsymbol{\Phi}_{m,i}}. \tag{5.13}$$

Here $\circ$ means an element-wise vector multiplication. By solving Eqs. 5.11 and using Eqs. (5.12) and (5.13), the gradient of the OF can be calculated.

5.3.6 Regularization

Tikhonov regularization is the most popular method to handle the ill-conditioning of inverse problem. It consists of applying an L^2 norm penalty term alongside the data fitting model [158]. The Tikhonov penalty term is a twice differentiable and smooth function that smooths out the edges in the reconstructed image. This is an adverse effect of Tikhonov regularization and in fluorescence imaging, as in other medical imaging modalities, it is desirable to use, a regularization term that preserve edges in the reconstructed image such as the total variation (TV) penalty [138]. For this reason, the TV based on the L^1 norm of the differences between neighboring voxels of the parameters to be reconstructed is implemented in the present work. Such a regularization term takes the form,

$$\mathcal{R} = \sum_{i,j=0}^{N_n, N_{nb}} \left\| \boldsymbol{\beta}_i - \boldsymbol{\beta}_j \right\|_1, \tag{5.14}$$

where i, j are the voxel indices corresponding to the volume voxels and neighboring voxels (6-connectivity is used here, *i.e.* nearest neighbors) respectively, N_n is the size of the vector of variables $\boldsymbol{\beta}$, and N_{nb} is the total neighboring pixel pairs ($N_{nb} = 6$). This form of the TV regularization enforces sparsity on pixel differences, and consequently tends to generate images with piecewise constant regions and sharp boundaries. This regularization function is not differentiable and to avoid this, it is modified to the form:

$$\mathcal{R} = \sqrt{v^T v + \delta_{TV}}, \qquad v = L\boldsymbol{\beta}, \tag{5.15}$$

where v calculates the nearest-neighbor differences via matrix L with dimensions $N_{nb} \times N_n$, and δ_{TV} is a small parameter that keeps the function near the true L_1 norm, while making it differentiable (machine precision is typically used; here the constant value 10^{-9} is used) [47].

5.4 Combined CW/TD Reconstruction

The central idea in this work is to use full-curve TPSF information in reconstructing images of the FLT while making the whole image reconstruction process computationally tractable. Two of the main tasks in FLT image reconstruction are the calculation and storage of the fluorescence source (Eq. (5.2)) for each node of the numerical model. These represent computationally and memory-wise intensive tasks. However, fluorophore distributions of interest in practice are generally sparse. Exploiting this fact, a combined CW/TD reconstruction scheme is developed which uses a fluorescence yield map obtained from a CW reconstruction as a mask for the full-curve TD reconstruction of the FLT. First, the measured TPSFs are smoothed out to filter out noise. Then, a CW reconstruction provide a fluorescence yield mask to fed to next step (TD reconstruction). The provided fluorescence yield mask with an initialized FLT map are fed to TD reconstruction step. A schematic of this scheme along with its different steps and a sample TPSF before each step is depicted in Fig. 5.1.

5.4.1 CW Step

Solving the forward problem in CW is much less computationally demanding in comparison with the TD case. The CW computation time is equal to that of one time step in TD, which makes CW reconstruction much faster TD reconstruction. The formulation of the CW image reconstruction problem is similar to that of the TD in that the time-dependent terms in Eqs. (5.1). The CW versions of Eqs. (5.2) and 5.3 are as follows:

$$\boldsymbol{Q}_m(\mathbf{r}) = \eta\mu_a^{x\rightarrow m}(\mathbf{r})\boldsymbol{\Phi}_x(\mathbf{r}), \tag{5.16}$$

$$A_{x,m}(\boldsymbol{\beta})\boldsymbol{\Phi}_{x,m} = \boldsymbol{Q}_{x,m}. \tag{5.17}$$

Note that rewriting all equations for the CW reconstruction problem is avoided here for saving space, and since these equations are very similar to those of the TD problem. One notable difference is that in the CW case there are only two adjoint variables, one for the excitation equation and one for the fluorescence equation.

In the inverse problem, the energy of measured TPSFs curve is used as CW measurements. The forward problem and inverse problem are solved in CW scheme.

5.4.2 Mask Producing Step

After CW reconstruction, the fluorescence yield distribution is thresholded to provide a mask for TD reconstruction. The threshold value used here is 0.25 of the maximum value of the reconstructed fluorescence yield. Node values in the discretization mesh for which the yield is smaller than the threshold are set to 0, and the other nodes retain they yield values as shown in the bottom right panel in Figure 5.1.

5.4.3 TD Step

Reconstruction of the FLT map is carried out only within active regions of the mask defining regions of interest (ROIs). More specifically, this means the fluorescence source is assumed to be zero where the mask is zero, and elsewhere (in ROIs) the discretized version of Eq. (5.2) is used to compute the source $\mathbf{Q}_m$. Then the calculated fluorescence source is used in the emission forward equation to find the fluorescence field at the boundary. This makes TD reconstruction much less computationally demanding while the full information of TPSFs is exploited in ROIs. Full-curve TD reconstruction algorithms have the potential for simultaneously reconstructing the FLTs for several fluorophores. This will be demonstrated in next section.

5.5 Results

Validation of the proposed combined CW/TD reconstruction scheme with numerical phantoms will now be described. First, a cylindrical phantom with 4 different FLTs is constructed to show the strength of full-curve TD reconstruction for distinguishing several lifetimes simultaneously. Second, the reconstruction algorithm is tested for two cases (2 and 3 lifetimes) on a complex geometry consisting of a numerical mouse phantom.

5.5.1 Noise Model

Poisson noise is added to the synthetic data to obtain realistic data (Fig. 5.2). Such noise is found in time-correlated single photon counting [110], which is widely used to acquire TD data. In a Poisson distribution, if N is the number of counted photons in a defined time interval, then the standard deviation is $\sqrt{N}$, and $SNR = \frac{N}{\sqrt{N}} = \sqrt{N}$. In TD-DOT, the photon counting time cannot take prohibitively long because it increases the total scanning time. In this work, the maximum number of photon is taken as $N = 1000$, corresponding to an SNR of $\approx$ 15 dB ($10 \log \sqrt{N}$), see Fig. 5.2; this is typical in practice.

5.5.2 TPSF Curve Smoothing

Since the image reconstruction algorithm uses each time bin of measured TPSFs, noise in the TPSFs can induce errors in the function/gradient evaluation. To suppress the noise, prior filtering of TPSFs is carried out using a Savitzky-Golay filter, which is a popular filtering method in digital signal processing for smoothing out noisy data [133, 142]. An output of this filter for a noisy TPSF is shown in Fig. 5.2

5.5.3 Simulation for 4 FLTs

We reconstruct the FLT map of a cylindrical numerical phantom containing 4 inclusions at different positions, each with the same fluorescence yield (0.05 mm^{-1}), but a different FLT (values of 0.4 ns, 0.6 ns, 1.0 ns, and 1.2 ns). Simulated measurements are acquired on 3 slices at the middle of the volume (inclusions are positioned at the middle of the cylindrical phantom - Fig. 5.3 left). The fluorescence data used for reconstruction comes from 12

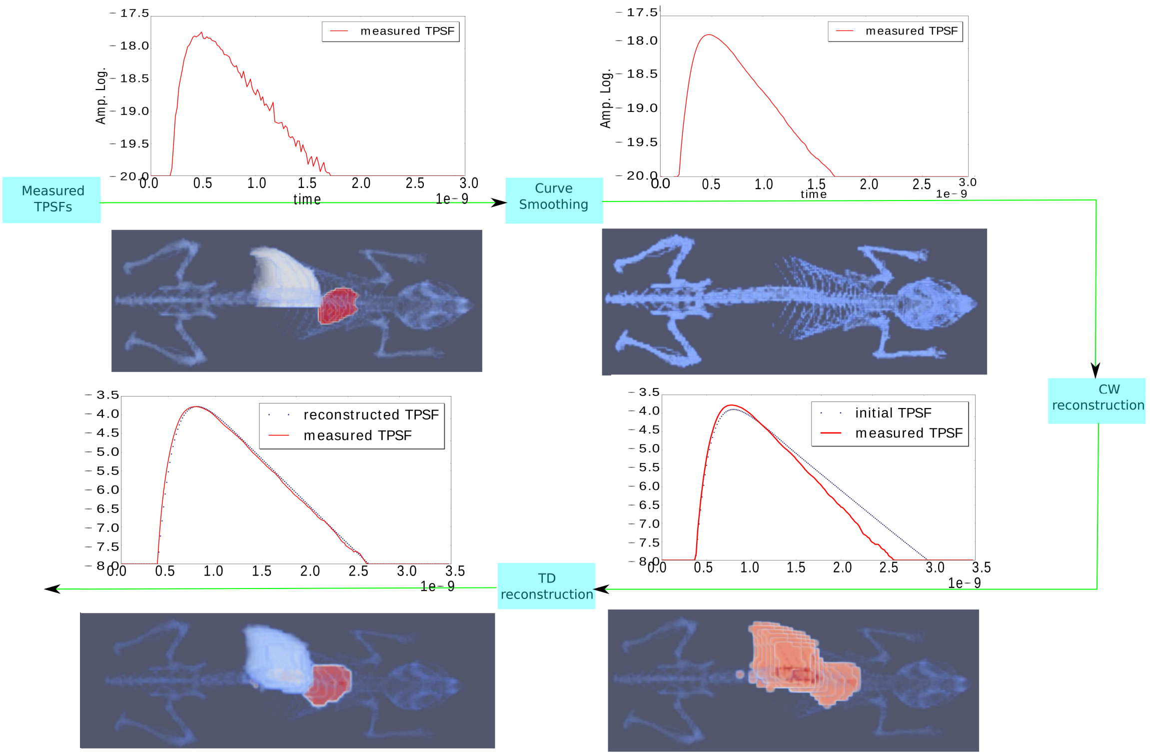


Figure 5.1 **Reconstruction steps and related TPSFs. Top left images show a sample measured TPSF and the target mouse image with labeled organs. Top right images show a smoothed TPSF and the initial map prior to the CW reconstruction step. Bottom right images show an initial TPSF with the initial FLT value and the fluorescence yield mask obtained from thresholding the CW yield image. This mask is fed to TD reconstruction step. Bottom left images show the fittted TSPF and the reconstructed FLT map after TD reconstruction.**

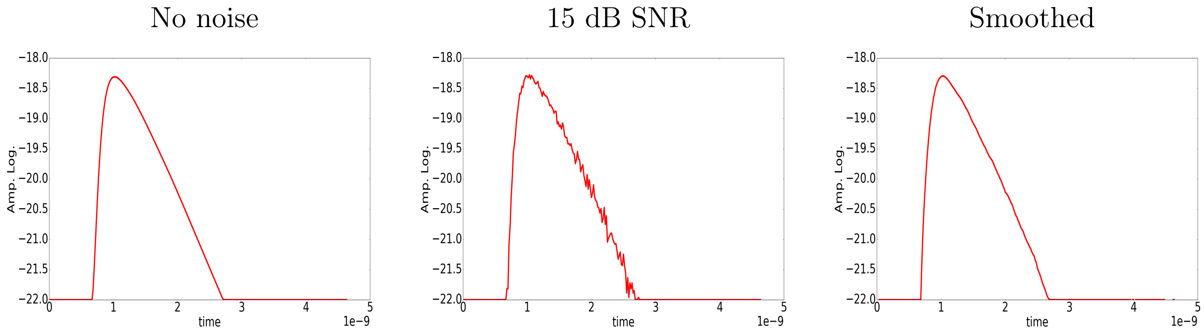

No noise
15 dB SNR
Smoothed

Figure 5.2 **TPSF model with noise and smoothed.**

projections for each slice, and 12 detection locations from 90° to 270° on the boundary for each projection (36 projections and 432 detection locations in total). Figs. 5.3 and 5.4 depict the reconstructed images and the histogram of the recovered FLTs.

As shown for a 4 mm distance between the inclusions, the reconstructed map allows distinguishing well the inclusions and their FLTs, and the histogram of the reconstructed FLT values shows that these are close to the real values. Also, there is a gap in the histogram graph between 0.6 ns and 1.0 ns as there should be, owing to the relatively large difference in FLTs. For 2 mm and 0 mm distances, one can still distinguish inclusions and their different FLTs, but the gap in the boundaries of the inclusions vanishes, and there is some cross-talk in the values of the FLTs (especially for the 0 mm distance). These are nevertheless excellent results considering the diffusive nature of light propagation the small distances between the inclusions and the fact that the FLTs can still be distinguished.

5.5.4 Numerical Mouse Phantom

To validate the proposed FLT reconstruction scheme for more complex geometries, synthetic data was generated using a numerical mouse phantom adapted from Digimouse [44], a numerical mouse phantom widely used in small animal medical imaging research. The mouse phantom was discretized to a structured grid with voxel size of 1 mm × 1 mm × 1 mm for a total of 22,340 nodes. Synthetic data was generated on a structured grid with voxel size of 0.75 mm × 0.75 mm × 0.75 mm to avoid the inverse crime. The optical properties of the different tissues contained in the mouse are obtained from [53]. Two experiments are defined to evaluate the performance of the proposed algorithm. For the first experiment, the heart and the left part of the liver are labeled by markers with the same fluorescence yield (0.05 mm^{-1}) and different lifetimes (0.4 ns and 0.6 ns respectively). In the second experiment, another FLT is added to the left lung with same fluorescence yield (0.05 mm^{-1}). The FLTs for the second experiment are 0.4 ns, 0.6 ns and 0.8 ns respectively for the heart, the left lung, and the left part of the liver of mouse phantom respectively.

In the forward model, Eq. (5.1) is solved for a time range of 3.0 ns and number of time bins $N_t = 180$ corresponding to $\Delta t \sim 16$ ps. Data is considered at mid-body of the mouse at longitudinal positions (z coordinate) separated by $\Delta z = 2$ mm over 20 mm along the length of the mouse. For each z position, the mouse is illuminated at 6 different angles differing by 60°. Fluorescence signals come from 21 detection locations in the transmission regime at angles separated by 4° ranging from 120° to 240° with respect to the illumination direction. This results in a total of 60 tomographic projections and 1260

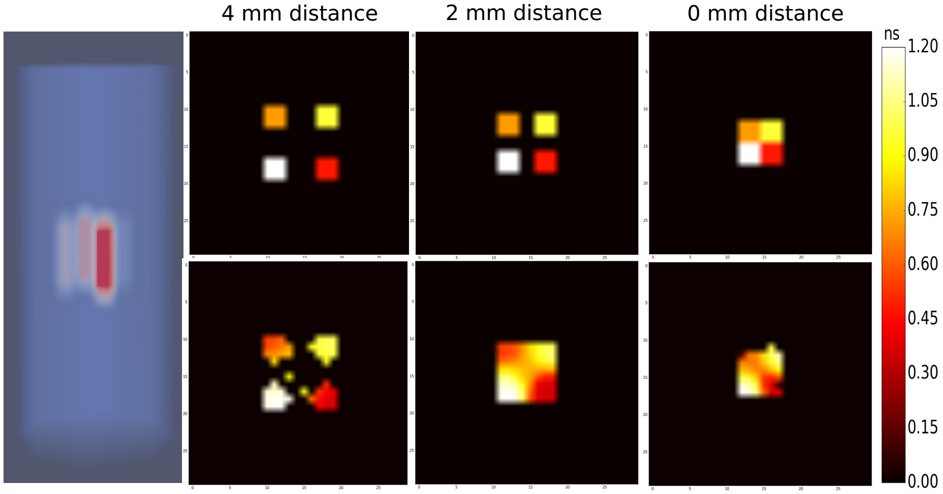


Figure 5.3 ***Reconstructed FLT map for 4 inclusions with different FLTs and spacing. First row and second row display a 2D slice of target phantom and reconstructed volume.***

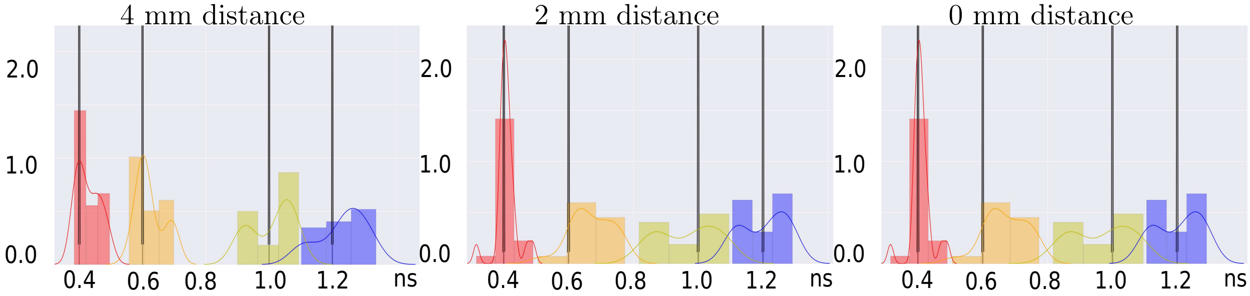


Figure 5.4 ***Histogram of 4 reconstructed FLTs with lifetimes*** $\tau = 0.4$ ***ns,*** $\tau = 0.6$ ***ns,*** $\tau = 1.0$ ***ns and*** $\tau = 1.2$ ***ns.***

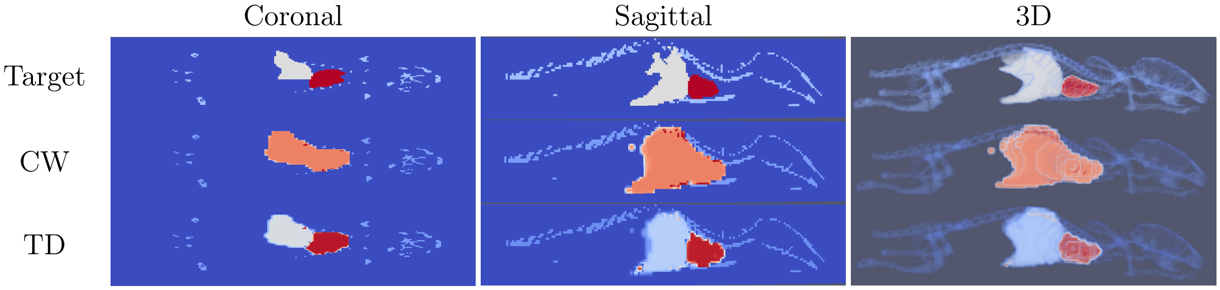


Figure 5.5 ***The target distribution and reconstructed distributions from CW and TD information are shown for lifetimes*** $\tau = 0.4$ ***ns in white and*** $\tau = 0.6$ ***ns in red.***

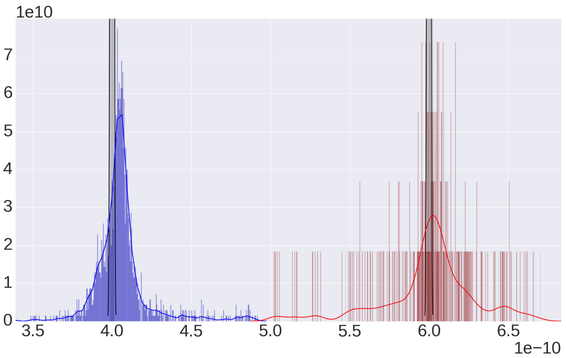


Figure 5.6 ***Histogram of reconstructed lifetimes for the mouse phantom with lifetimes*** $\tau = 0.4$ ***ns and*** $\tau = 0.6$ ***ns.***

TPSFs being acquired. The results of the first experiment (2 FLTs) for sagittal, coronal, and 3D views are displayed on Fig. 5.5.

One can see that the fluorophores with distinct FLTs are well localized in the TD reconstructed images. Additionally, the reconstructed FLTS are seen to be very close to the real values as shown in the histogram of Fig. 5.6. The results of the second experiment (3 FLTs) is depicted in Fig. 5.7. Reconstructed FLTs for each organ are distinguishable. Also the histogram of the reconstructed FLTs are close to real value as expected. This is an excellent result considering that the different FLTs are in tightly packed organs. The combined CW/TD reconstruction carried out in 3D took almost 8 hours of computing time on a standard desktop computer (Intel Core i7-2700K CPU 3.50 GHz × 4 and 32.00 GB RAM). Programming was done in Python 2.7 (Numpy/Scipy packages [81, 159]) and Cython.

5.6 Conclusion

Full-curve TD reconstruction uses each time bin of TPSFs. This makes the whole iterative image reconstruction process computationally very demanding. In this work, a combined CW/TD model-based image reconstruction scheme for FLT tomography is proposed, which significantly reduces the computational burden, making image reconstruction practical, as it concentrates on regions of interest based on a fluorescence yield mask obtained with a CW data-based reconstruction. This allows obtaining images in accessible time on a standard desktop computer. This is a major leap forward in full-curve TD fluorescence tomography. Since in the reconstruction process the values of TPSFs in each time bin are used. In such a reconstruction scheme, noise can easily propagate in the function and gradient evaluation. This can lead to a wrong descent direction in the optimization algorithm. To suppress the noise a smoothing filter is applied to each TPSF prior to image reconstruction. The smoothed TPSFs are then normalized to their energies. TPSF normalization has a great impact in practice since it eliminates difficult to calibrate experimental factor such as optical and electrical efficiency and gain of instruments. In the reconstruction process, first a CW objective function is defined for the fluorescence yield distribution, which is then minimized with an L-BFGS algorithm. The output of this step is thresholded and fed as a mask to a full-curve TD image reconstruction to recover the fluorescence lifetime map. One bottleneck of full-curve TD image reconstruction is solving Eq. (5.2) for each node of the spatial discretization grid. Solving the forward model for nodes sorted out by the mask accelerates image reconstruction. Reconstruction of the fluorescence yield using CW information (energy of TPSFs) and of the FLT with a mask from the yield image, make the whole process less computationally demanding, while the whole information of TPSFs is meaningfully used. The proposed algorithm is validated with two numerical experiments. With the use of full-curve information in the reconstruction process, it has demonstrated its ability to reconstruct more than 2 different FLTs with a great contrast, which can be found in the literatures. In this work, we reconstructed for 4 different FLTs to show the strength of full-curve TPSF reconstruction. The results show a good distinction between inclusions with different FLTs, even when separated by a small distance. To demonstrate it ability to handle complex geometries, a numerical mouse phantom is used. Two experiments were conducted on the torso of the mouse, where the organs are tightly positioned with respect to each other, making this a challenging imaging problem, especially in diffuse optical imaging. First, two different FLTs were ascribed to in the heart and part of the liver (left side) of the mouse. In the reconstructed image, a sharp edge between the two organs allows distinguishing them clearly with high contrast. Second, another FLT is added to the left lung. Reconstructed results

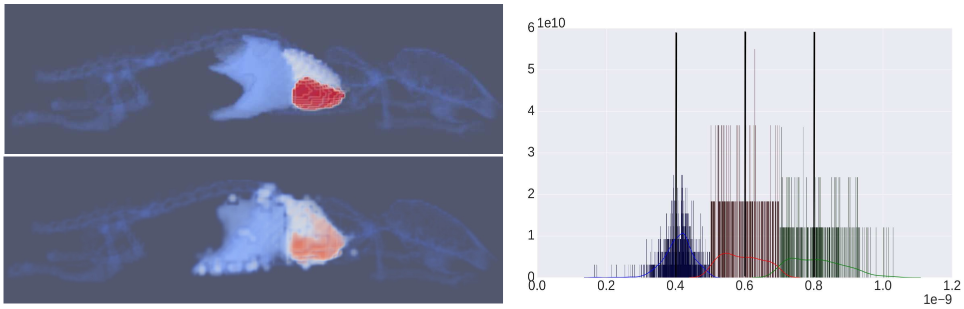


Figure 5.7 ***3D view and histogram of reconstructed FLTs for the mouse phantom with lifetimes τ =0.4 ns, τ =0.6 ns and τ =0.8 ns.***

show a good contrast between organs with different FLTs. The whole reconstruction time was about 8 hours for the mouse phantom. This time could potentially be further reduced with more powerful computing hardware and further code optimization.

CHAPTER 6

MACHINE LEARNING-BASED IMAGE RECONSTRUCTION

6.1 Preface

Title:
A Deep Learning Generative Model for Fluorescence Diffuse Optical Tomography Image Reconstruction

Authors and affiliation:

- SeyedRohollah Ettehadi: PhD student, Université de Sherbrooke, Faculté de génie, Département de génie électrique et de génie informatique, Laboratoire TomOptUS.
- Y. Bérubé-Lauzière : Professeur (directeur de doctorat), Université de Sherbrooke, Faculté de génie, Département de génie électrique et de génie informatique, Laboratoire TomOptUS.

Submission date:
21 December 2017

Journal:
IEEE Transaction on Medical Imaging

Contribution of the document to the thesis

This article presents an alternative image reconstruction algorithm for iterative model-based image reconstruction algorithms, which is widely used in the medical image reconstruction and specially in the field of diffuse optical tomography. The new image reconstruction algorithm is developed based on the deep learning techniques. The proposed deep learning-based image reconstruction model can be applied to others imaging systems since it is independent from data acquisition strategy.

Abstract:

In this study, a deep learning-based image reconstruction algorithm for fluorescence diffuse optical tomography is proposed. Among all deep learning models, the generative models are a promising choice for image reconstruction. These models are able to generate high-dimensional data such as an image from a low-dimensional latent space. A deep encoder network is used to learn a joint manifold low-dimensional latent space from given dataset. One of the most successful generative models for manifold learning is the variational autoencoder. The proposed image reconstruction algorithm uses such an autoencoder model to learn a low-dimensional joint manifold between the sensor and image domains. First, a forward model generates a bulk of continuous-wave fluorescence synthetic measurement data. Then the proposed model is trained with this synthetic data to recover the fluorescence distribution. The trained model shows clear and promising results form sparse measurements when it is applied to a real phantom. Also, to evaluate the performance of the proposed algorithm, it is trained and validated on numerical phantoms with different sizes.

6.2 Introduction

Tomographic medical image reconstruction has seen spectacular development in the last decades thanks to steady improvements in hardware and algorithms. All medical imaging modalities such as magnetic resonance imaging (MRI) [51], X-ray computed tomography (CT) [82], positron emission tomography (PET) [86], ultrasound [154] and diffuse optical tomography [11] use direct/iterative algorithms based on state-of-the-art numerical techniques and computing resources to recover/reconstruct the interior structure/properties of an object. In an image reconstruction process, an inverse function directly or iteratively transforms the measured data from the sensor domain to an image. During data acquisition, an intermediate representation of the object to be reconstructed is encoded in the sensor domain. The inverse transfer function plays the role of a decoder that allows visualizing the object represented in the sensor domain. Since the encoding step is not exact owing to the non-idealities of sensors and noise, in the decoding step the inverse function approximates the image of the object and often requires the details of the data acquisition approach to tune and optimize the reconstruction performance.

Deep neural network models have allowed breakthroughs in the last few years [143]. They have drawn a lot of attention in many applications. In general, deep learning is a data analysis framework with several layers that promote a higher level of abstraction and prediction from data [18, 143]. Deep learning resorts to deeper artificial neural networks that enable analysis models to account for more non-linearity and complexity through the many layers used. Each layer consists of many neurons that are inter-layer connected. In the training process, a large set of unlabeled and labeled data are fed to the input/output layers of a network, and weights of the neurons are updated along the gradient descent direction of a loss function. The important factors that enable these algorithms to achieve exceptional performance are a bulk of dataset (big data), the availability of high-performance computing power, smarter weight initialization, and significantly deeper network architectures.

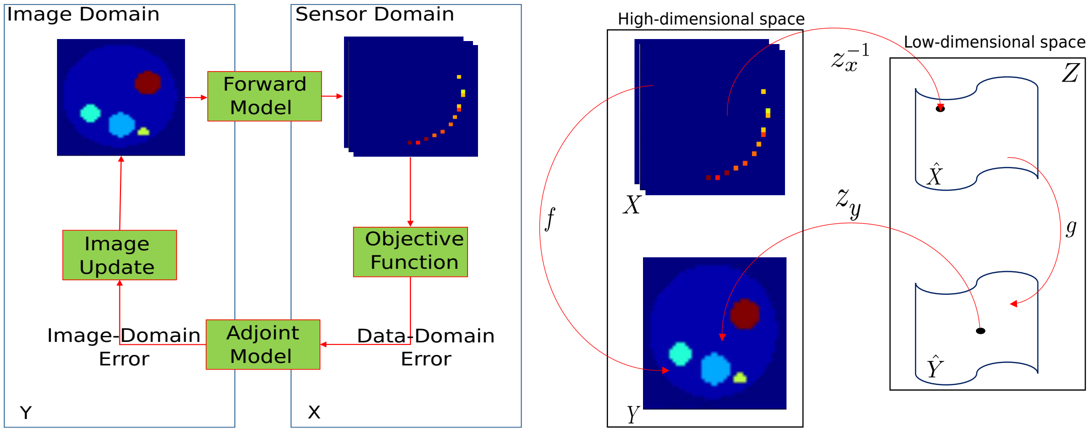
The performance of deep networks is now well reported in the areas of computer vision [49], speech recognition [72], and text/image processing [36, 169]. Machine learning and more recently deep learning techniques have been applied in medical imaging analysis/processing for lesion detection [147, 149, 160], segmentation, and classification [16, 27, 56, 130, 140]. In the medical image reconstruction field a few approaches have been applied for improving low-dose X-ray CT images [83], transformation of k-space to image space in MRI [171], and enhancement of photo-acoustic reconstructed images [5]. There is yet still relatively little work on the application of deep learning to medical image reconstruction [162]. Here we propose for the first time an algorithm based on deep learning to directly reconstruct tomographically an image of the distribution of fluorescence sources inside biological tissues, more precisely of the fluorescence yield of such sources, from the power of measured fluorescence signals measured at the boundary of such media.

This paper is organized as follows: Section 6.3 provides background on tomographic optical imaging in biological tissues. Section 6.4 discusses deep learning-based image reconstruction and associated requirements pertaining to the present work. Section 6.5 provides the details of data preparation and the deep generative model used herein. Phantom experimental details along with results are described in Section 6.6. Finally, Section 6.7 concludes the paper.

6.3 Background and Motivation

Diffuse optical tomography (DOT) aims at imaging non-invasively the spatially varying optical properties (absorption and scattering coefficients) inside a biological tissue [22, 67], to be called the "medium" in the sequel. This leads to an inherently ill-posed inverse problem because of the diffusive nature of light propagation in such media. In fluorescence DOT (FDOT), an external source of light, typically a laser, excites a fluorescent compound injected into a biological tissue, and then the emitted signal is detected. These signals are then used to image in 3D the distribution of the fluorescent compound or some of its properties such as its fluorescence lifetime [67, 104]. For example, one may localize cancer when the compound is designed to label cancer cell receptors, or specific proteins related to cancer [125].

Most of the current FDOT reconstruction algorithms resort to an iterative model-based image reconstruction (MBIR) scheme [6, 90]. In these algorithms, first, a light propagation model predicts the light distribution in the medium under examination along with the measurements on its boundary at a set of detector locations. Then, a data fitting function (objective function) is used to estimate the discrepancy between experimental measurements and predictions thereof. The objective function is minimized through an iterative optimization algorithm (*e.g.* gradient-based). It must be emphasized that due to the ill-posedness and the nonlinearity of the inverse problem, reconstructing the optical properties of tissues or the distribution of a fluorescent compound is a challenging task [41]. These methods suffer in image quality performance when the number of measurements is limited or in the presence of noise, which are typical in FDOT image reconstruction. Moreover, these algorithms are time consuming because of their iterative nature. This makes such algorithms unsuitable for real-time optical imaging applications.



(a) Iterative algorithms (b) Joint manifold learning

Figure 6.1 ***Flow charts of image reconstruction for (a) iterative algorithms and (b) joint manifold learning.***

To overcome these limitations in FDOT imaging, a deep-learning-based image reconstruction algorithm is herein proposed. To the best knowledge of the authors, this is the first deep learning algorithm for FDOT image reconstruction. The algorithm can reconstruct an image of a medium in less than a second, thus making it appropriate for real-time image reconstruction. Learning algorithms require an enormous training dataset to perform well. First, a light propagation model provides a bulk of simulated data for training and validation. Then the deep neural network is trained and validated with the simulated dataset. Finally, the deep network is tested on real data measured in the laboratory on a phantom mimicking a real optical biomedical imaging situation to evaluate its image reconstruction performance.

6.4 Theory and Method

In FDOT image reconstruction, the fluorescence spatial distribution is reconstructed from a set of measurements on the boundary of the medium. This means the encoded information in the sensor domain must be decoded to a spatial map (image). Deep generative models consisting of encoder/decoder neural networks are well-suited for this purpose, as these models are capable of generating high-dimensional images from sampling a much lower dimensional space.

One of the well-known deep generative models is the variational autoencoder (VAE). Autoencoders are networks with the same input/output size. First, an encoder network infers a low-dimensional dense representation of the set of inputs (simply called the input) by way of so-called latent variables. For this reason, an encoder network is also called an inference network. The output of the encoder can be interpreted as a manifold of inputs (the input manifold). Then a decoder reconstructs the set of outputs (simply called the output) from the latent variables. In variational autoencoders, instead of encoding the input to an explicit vector in the latent space, a density function (normally unit Gaussian) of each set of inputs is encoded in the latent space [88]. This encourages the model to place a high probability mass on many latent variables that could have generated samples, rather than collapsing to a single point estimate of the most likely value [60]. This Gaussian distribution causes the model to be less sensitive to small variations in the input such as noise or other artifacts that may affect measurements during data acquisition. This turns the autoencoder into a probabilistic generative model [43]. VAEs are simple to implement, and can be optimized with an element-wise error between input and output, *e.g.* square error (loss function).

A VAE can learn a predictable coordinate system when the encoder network is simultaneously trained in combination with the decoder network, which makes a VAE an excellent manifold learning algorithm [60]. This is very interesting for the present application, since an image is to be reconstructed by sampling from a predictable coordinate system (low-dimensional manifold space). Consider an input (measurements here) $x \in X$ and an output (reconstructed image here) $y \in Y$ in a high-dimension space $X \times Y$ and manifolds of them $\hat{x} = z_x^{-1}(x), \hat{y} = z_y^{-1}(y)$ in a low-dimensional space $\hat{X} \times \hat{Y}$ (Fig. 6.1(b)). Then the VAE model approximates the function f by finding a projection g from the low-dimensional input manifold to the low-dimensional output manifold:

$$f \approx \hat{f} = z_y \circ g \circ z_x^{-1}. \tag{6.1}$$

6.4.1 Conditional Variational Autoencoder

The conditional VAE (CVAE) is a variant of the VAE in which the probability distribution of a dataset is conditioned by another distribution. In the imaging problem considered here, this is translated in the image distribution being conditioned by the measurement distribution. Such conditioning forces the model to reconstruct the image by sampling from a conditional latent space (joint manifold space). In the model proposed here, the CVAE is used to reconstruct the fluorescence map by sampling from the learned joint manifold space. The measurements are the input to the model and the reconstructed image is the output. The details about how this is implemented will be discussed later in Sect. 6.5.3. In the CVAE used here, the encoder network maps the input from a high-dimensional sample data distribution $x \in X$ (measurement domain) to a lower-dimension continuous latent space Z (joint-manifold coordinate). The decoder's role is to invert this process and generate data $y \in Y$ (fluorescence map) by taking a sample from the latent space. Since Gaussian noise is added to the input ($x \to x'$), the proposed algorithm is similar to a denoising VAE [75]. The difference between a denoising VAE and a VAE is the noise which is injected into the input of the model to improve the VAE's performance. In fact, learning from a distribution affected by noise heuristically leads to a model that can handle a broader set of inputs, thus making it more robust. Here, to enable this the CVAE [43] and denoising formulations [75] are combined. This leads to writing the loss function as the sum of the reconstruction mean-square error (MSE) and a prior regularization term as follows:

$$\mathcal{L}_{CVAE}(\phi, \theta; x', y) = \mathcal{L}_{MSE} + \mathcal{L}_{prior}, \tag{6.2}$$

where

$$\begin{aligned} \mathcal{L}_{MSE} &= -\mathbb{E}_{p(x'|x)}\mathbb{E}_{q(z|x',y)}[\log p_\theta(y|z,x)], \\ \mathcal{L}_{prior} &= \mathbb{E}_{p(x'|x)}[D_{KL}(\hat{q}_\phi(z|x',y)||p(z|x))], \\ \hat{q}_\phi(z|x',y) &= \mathbb{E}_{p(x'|x)}[q_\phi(z|x',y)]. \end{aligned} \tag{6.3}$$

Here $q_\phi(.)$ and $p_\theta(.)$ are the parametrized distributions of the encoder and decoder networks with ϕ and θ being parameters to be optimized during training, $p(z|x)$ is a sampling from a unit normal distribution conditioned by the distribution of x ($z \sim \mathcal{N}(0, I)$), and D_{KL} is the Kullback-Leibler divergence between the approximated inference of the conditioned encoder network and the probability of the conditioned decoder network.

The $\mathcal{L}_{MSE}$ term in Eq. (6.3) is the log-likelihood of the variables ϕ, θ, x, x' and y under the approximate posterior over the latent variables. It is also called the reconstruction error, and is equivalent to a least-square error function between the image generated by the network ($\hat{f}$ function) and the true distribution y ($\mathcal{L}_{MSE} = \left\| y - \hat{f}(\phi, \theta; x, x') \right\|_2^2$; see Eq. (6.1) and recall that $y = f(x)$). $\mathcal{L}_{prior}$ as it appears in Eq. (6.3) is the entropy of the approximate posterior. It minimizes the expectation of D_{KL} between the true posterior distribution $p(z|x)$ and the approximate posterior distribution $\hat{q}_\phi(z|x', y)$ over all noisy inputs sampled from $p(x'|x)$. In fact, this term serves the purpose of approximating the inference model by the probability of the prior when they approach each other. The right hand side of Eq. (6.12) can be optimized via stochastic gradient descent. At reconstruction time sampling from the distribution $p(y|x')$ can be simply replaced by sampling z from $\mathcal{N}(0, I)$. Note that the proposed model is implemented in the TensorFlow [1] framework, developed by Google for deep learning applications.

6.5 Training Details

6.5.1 Synthetic Data Generation

A light propagation model, so-called the forward model, is used to predict the light flux at detector locations on the surface of a medium from a given distribution of the optical properties and fluorescent compounds inside the medium [50].

The general model for light propagation inside biological tissues is the radiative transfer equation (RTE). However, owing to the complexity of solving the RTE, the diffusion equation (DE) is ubiquitous in biomedical optics. The DE is a partial differential equation (PDE) derived from the RTE using the diffusion approximation [164] which assumes that the light field propagating in the medium is isotropic at each point.

Herein, the DE is used as the forward model. It is solved by finite differences that spatially discretizes the light propagation equations. To obtain a training dataset, 50,000 measurements are simulated using different meshes varying from 1.5 mm to 2.0 mm in steps of 0.05 mm (5,000 measurements for each grid size). Different mesh sizes are used in order not to bias the dataset by a specific mesh size. Similarly, for the validation dataset, 5,000 measurements are simulated using mesh sizes from 1.0 mm to 1.5 mm (1,000 measurements for each mesh size). One to 4 cylindrical inclusions with random diameters varying from 2 mm to 6 mm are randomly located inside a numerical cylindrical phantom with a diameter of 25 mm (same dimension as the real phantom used in the experimental measurements). To promote a certain degree of robustness to translations during training, the image dataset is translated randomly 1 to 2 pixels up, down, left, and right. Also, Gaussian noise with a power of 1% to 10% of the generated signal is added to that signal (corresponding to SNRs from 20 dB down to 10 dB). This is a typical range in optical measurements. This added noise turns the model into a denoising model and promotes noise robustness and convergence during training [75].

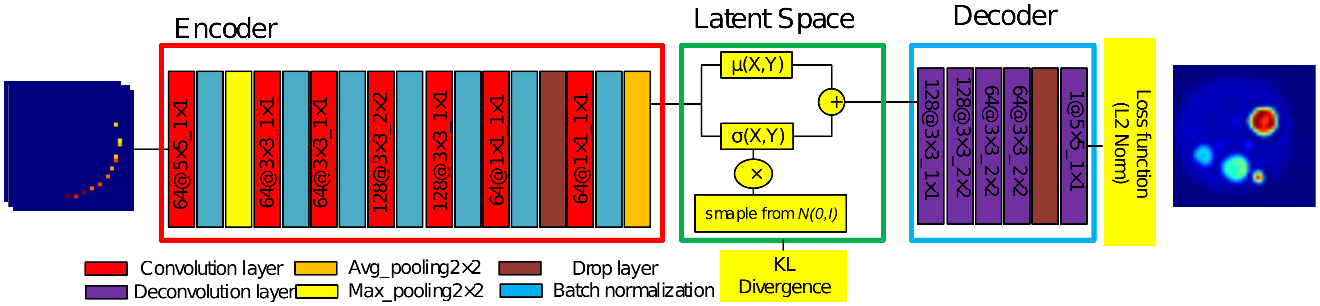


Figure 6.2 ***Network architecture of the implemented VAE with CNN.***

6.5.2 Preprocessing Step - Scaling

Before feeding data to a neural network, a preprocessing step is necessary to scale the data in the training dataset to the range $(-1, 1)$. This is very important as no scaling can lead to failure in convergence and learning. Such scaling is in fact a general requirement when gradient-based optimizers are used. In diffuse optical imaging as considered here, since the light intensity typically decays exponentially in media, simulated measurements are first converted to a log scale. Then the mean is subtracted from the data and the resulting data are divided by their maximum value. The same scaling is performed on the fluorescence distribution map that is used as the ground-truth image for training.

6.5.3 Network Architecture

Fig. 6.2 depicts the architecture of the network implemented in this work. The network is designed to operate on 2D transverse images (axial slices) of the object to be reconstructed, each slice being of size 32×32. All of the input images are resized to the same size before being fed to the network, and the output images are also of this size (autoencoder networks operate under such settings as mentioned previously).

The input is formatted so as to incorporate the shape of the medium along with the detection locations of each measurement (detection locations are mapped in this 2D image, see leftmost set of images sketched in Fig. 6.2). The input is a set of 2D images with several channels. Each channel corresponds to a tomographic projection (set of measurements for a given source position), and the number of channels is equal to the number of projections for each slice. The representation of a given channel is a sparse 2D image with zero values everywhere except at detection positions where it takes the value given by the detector. The number of channels can be extended with other *a priori* information that one may have about the medium such as optical properties (absorption and scattering), the shape of the medium, and refractive index. Here an extra channel was added, consisting of the shape of the medium at the slice considered. This extra channel informs the model about boundary information, which is very important when the model is trained for a medium with a different shape at each slice (here training is for a cylindrical medium, hence this extra channel is the same for all slices, but for the sake of generality, this information is nevertheless considered).

As shown in Fig. 6.2, the encoder is implemented as a convolutional neural network (CNN) consisting of several convolutional layers. Several VAE architectures were implemented to find the one best suited to our problem. The best performance for our dataset is obtained with the architecture proposed in [141] with small modifications. More precisely, the encoder uses 7 convolutional layers each with batch normalization and a leaky rectified linear unit as the activation function. In the decoder, 5 layers of deconvolution are used along with a drop layer with drop coefficient 0.5 for regularization. The details of the filter sizes and strides in the CVAE used here can be found in Tables 6.1 and 6.2.

Table 6.1 ***Architecture parameters of the encoder.***

	Filter size	Stride	Depth	Output
Conv1	5×5	2×2	64	$16 \times 16 \times 64$
Conv2	3×3	1×1	64	$16 \times 16 \times 64$
Conv3	3×3	1×1	64	$16 \times 16 \times 64$
Conv4	3×3	2×2	128	$8 \times 8 \times 128$
Conv5	3×3	1×1	128	$8 \times 8 \times 128$
Conv6	1×1	1×1	64	$8 \times 8 \times 64$
Conv7	1×1	2×2	64	$4 \times 4 \times 64$

Table 6.2 ***Architecture parameters of decoder model.***

	Filter size	Stride	Depth	Output
Deconv1	3×3	1×1	128	$4 \times 4 \times 128$
Deconv2	3×3	2×2	128	$8 \times 8 \times 128$
Deconv3	3×3	2×2	64	$16 \times 16 \times 64$
Deconv4	3×3	2×2	64	$32 \times 32 \times 64$
Deconv5	5×5	1×1	64	$32 \times 32 \times 1$

In the optimizing process, an ADAM [87] optimizer with minibatch size of 100, learning rate 0.0002, and decay rate 0.98 is used. As it is mentioned before, the loss function is a simple L_2 norm of error between the reconstructed image and the ground-truth image. For the regularization, drop layer with drop rate 0.5 is applied before the last convolutional layer in both encoder and decoder networks as it shown in Fig. 6.2.

6.6 Results

6.6.1 Real Phantom Experiment

For the first experiment, our model is trained for an optical phantom with known absorption and scattering coefficients (μ_a and μ_s' respectively) to reconstruct the fluorescence yield ($\eta\mu_a^{fl}$) from continuous-wave (CW) measurements. Training and validation datasets are generated with a size of 50,000 and 5,000 samples respectively as mentioned in Sect. 6.5.1, with the optical properties in Table 6.3.

Table 6.3 ***Optical parameters of the real and numerical phantoms.***

	Real phantom	Numerical phantom
μ_a (mm^{-1})	0.0231	$0.015 - 0.025$
μ_s' (mm^{-1})	1.57	$1.0 - 2.0$
n	1.51	$1.45 - 1.55$

Our group has developed a time-domain DOT scanner [110] that allows acquiring time-domain optical measurement data at a plurality of positions around an object. Here, the CW measurement at a given position is obtained by integrating the area under the TD measurement thereat (this integral corresponds to the light energy contained in the TD measurement).

The phantom used in this experiment (Biomimic$^{\text{TM}}$ solid optical phantom; $\mu_a = 0.0231\,\text{mm}^{-1}$ and $\mu_s' = 1.57\,\text{mm}^{-1}$, height $= 90$ mm, diameter $= 2.5$ cm, INO, Québec City, QC, Canada) is made of a material mimicking the optical properties of biological tissues. It has two bores with diameters of 3 mm and 5 mm. These bores were filled with indocyanine green (ICG) at a concentration of 10 μM. ICG is a fluorescence dye commonly used in medical diagnostics [4].

The measurements are acquired around the phantom at different heights between 30 mm and 70 mm at step of 2 mm along the phantom (20 slices in total). Twelve source locations (projections) are used to illuminate the phantom at each height with an angle of 30° between each source location. Twelve detection locations collect the fluorescence signal from 120° to 240° with respect to the illumination direction (transillumination regime) at step of 10°. The total number of measurements acquired is $20 \times 12 \times 12$ which are mapped to a dataset of 2D images with the size $20 \times 32 \times 32 \times 12$.

To investigate the effect of sparse measurements on the proposed learning-based image reconstruction algorithm, the model is trained with different projection numbers: 3, 6, and 12 for each slice. The 3D reconstructed volume and 2D top view are displayed in Figs. 6.3 and 6.4 respectively. As can be seen, the model can recover the fluorescence yield distribution of ICG even with very sparse measurements provided by only 3 projections for each slice. Note that the reconstruction in this example is not quantitative since our scanner does not provide absolute calibrated measurements of light fluxes exiting the medium (it is a time-domain scanner and it is not designed for this). Hence, the fluorescence yield spatial distribution can be imaged with high accuracy, but its value cannot be recovered quantitatively. However, when calibrated measurements are available,

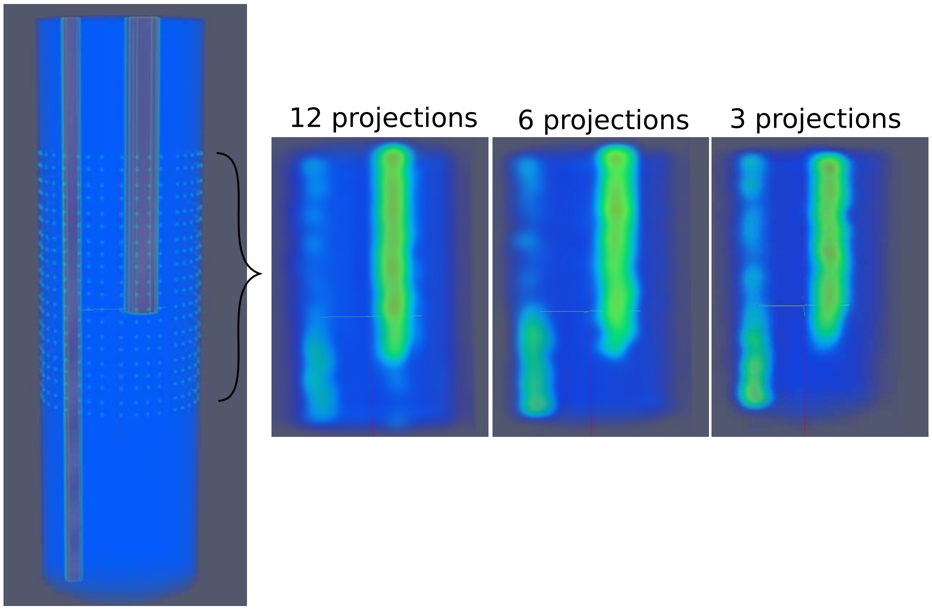


Figure 6.3 ***3D view of the phantom used (left) and of the reconstructed fluorescence yield distribution in the phantom for different number of projections.***

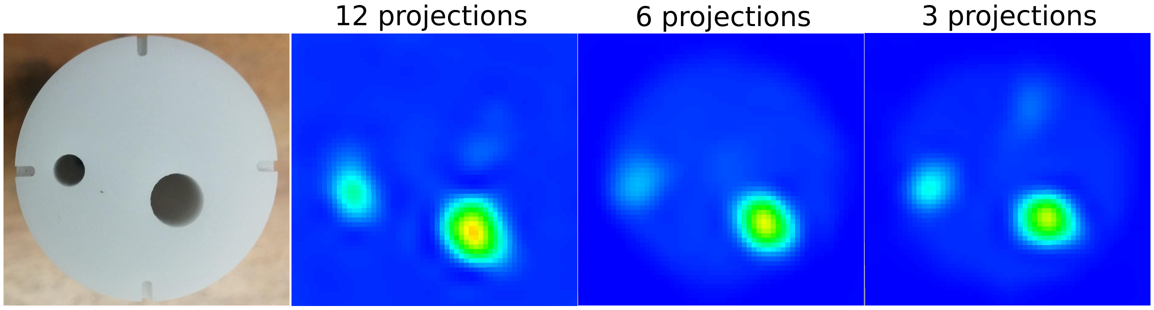


Figure 6.4 ***Top view of reconstructed phantom for different numbers of projections.***

as in the next example, then it is possible to recover a fluorescence yield image with good quantitative accuracy.

6.6.2 Cylindrical Volume with Different Sizes

To evaluate the capability of the proposed algorithm to generalize for volumes with different sizes, the model is trained with synthetic data for two cylindrical volumes with diameters of 25 mm and 50 mm, and then the algorithm is asked to reconstruct the fluorescence yield distribution for a cylindrical volume with a diameter of 35 mm.

As before, training datasets and validation datasets with a size of 50,000 samples and 5,000 samples respectively are generated for each volume (25 mm and 50 mm) totaling 100,000 training samples and 10,000 validation samples. For the test dataset, a series of measurements are generated from a cylindrical volume with a diameter of 35 mm. Two inclusions are placed at different positions as depicted in Fig. 6.5 (top row).

The reconstructed results shown in Fig. 6.5 (bottom row) indicate that the proposed model can recover the fluorescence distribution for a volume that is not considered in the training data. This shows that learning-based image reconstruction algorithms can reconstruct the fluorescent distribution for an unknown volume. This is impossible in model-based image reconstruction, whereby the exact geometry of the volume to be reconstructed must be provided beforehand to the algorithm. Furthermore, as the colorbar scale shows at the right in Fig. 6.5, the fluorescence yield is reconstructed with good quantitative accuracy. This is possible, since in this case the measurements contained in test dataset are quantitative by nature since they are synthetically generated through modeling. Quantitative optical measurements are also possible experimentally with imaging scanners. This requires calibrating the scanner in absoluted terms with light source standards.

6.7 Conclusion

In this paper, deep learning for FDOT image reconstruction was investigated. To the best of the knowledge of the authors, this is the first time a learning model is used in FDOT reconstruction. A deep generative model was proposed, which can learn an optimal inversion function between measurement and image space. The proposed model is adapted from the VAE model which is a well-known model for unsupervised learning and variational inference. VAEs consist of encoder and decoder networks. The encoder network maps the measurements from a high-dimensional space to a low-dimensional joint-manifold space.

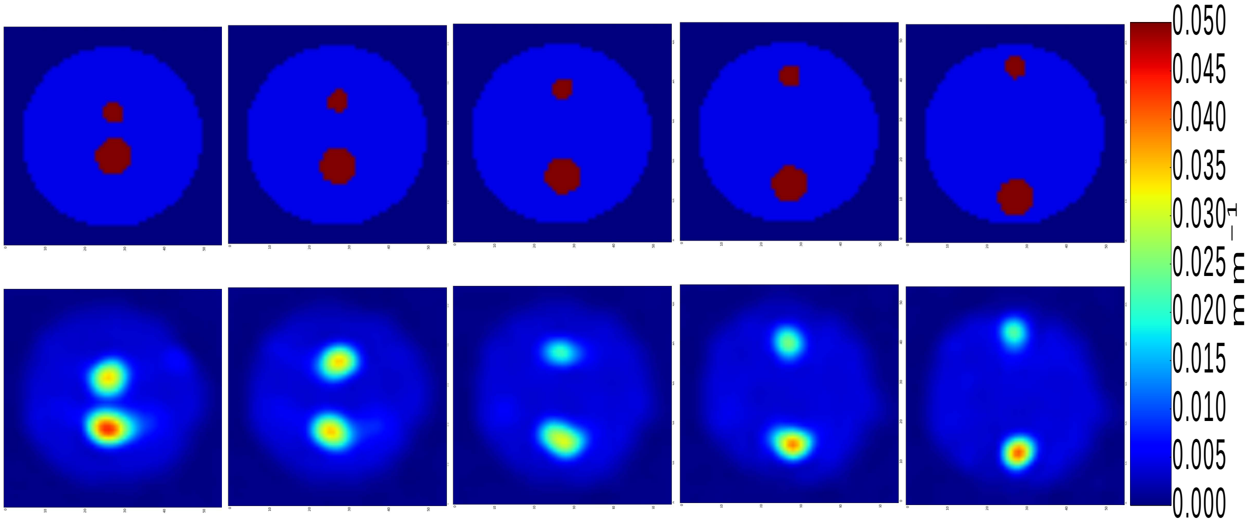


Figure 6.5 ***Reconstruction for a cylindrical volume with 35 mm diameter. Top row shows 2D slices of the phantom for 5 different positions of fluorescent inclusions. Bottom row shows the reconstructed corresponding fluorescence yield maps.***

By sampling from that space, the decoder can reconstruct an image. The model is first trained with a bulk of dataset which can be synthetically generated from a forward model (note that real data could also be used for that purpose). After training, it can reconstruct a slice image in less than 1 second. In the present case, the training time was around 8 hours. The proposed model thus shows great potential and paves the way for real-time diffuse optical tomographie imaging applications, which are not yet possible.

The proposed algorithm showed great robustness to noise and sparse measurements, which is also not possible with current image reconstruction techniques. First, the model was trained with synthetic data, and then was tested with experimental data. The recovered image features were well localized even for very sparse measurements. In the second experiment, the model's performance is evaluated for a more general problem. In this case, the model was trained for two different cylindrical volumes with different sizes. Then it was asked to reconstruct a volume with an unknown size between the two training volumes. This shows the capability of the model to generalize to objects of sizes other than those used for training. Such a capability is not possible with current standard image reconstruction techniques, whereby the geometry must be provided beforehand to the reconstruction algorithm.

Deep learning models, such as the model proposed here, can be potentially extended to other imaging modalities since the learned function is independent of the data acquisition strategy. One limitation for learning algorithms is preparing a big dataset for each medium with different sizes and shapes. In the medical imaging field, this limitation can be compensated by developing precise forward models to simulate measurements made on the medium during scanning to generate synthetic data for training deep learning models.

Acknowledgements

YBL acknowledges financial support from NSERC (Discovery Grant) for this work. YBL is member of the FRQ-S-funded Centre de recherche du Centre hospitalier universitaire de Sherbrooke (CR-CHUS) and of the Institut quantique at Université de Sherbrooke.

6.8 APPENDIX

In the variational autoencoder a values of z is sampled to be likely produced x, and computed $p(x)$ just from those. A new function $q(z|x)$ is approximated by encoder that takes a value of x and produce a distribution over z values that are likely to produce x (Fig. 6.6). The space of all z's are under a low-dimensional space $q(z)$ with the prior $p(z)$. Let's start with calculating $p(z|x)$ as below:

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} = \frac{p(x,z)}{p(x)}, \tag{6.4}$$

where $p(x) = \int p(x|z)p(z)dz$ is a marginal distribution and its computation is difficult specially for high-dimensional space. In variational inference we compute $p(z|x)$ by approximating it with another distribution $q(z)$ which is a low-dimensional space. This distribution is a tractable distribution such as Gaussian. It can be adjusted by tuning its parameters such as μ and $\sum$ to be closed to the $p(x|z)$. Therefore, it can be computed and produced values that are likely to $p(x|z)$. To make $q(z)$ close to $p(x|z)$, we minimize the Kullback-Leibler divergence as below:

$$KL\left(q(z)||p(z|x)\right) = -\sum q(z)log\frac{p(z|x)}{q(z)}. \tag{6.5}$$

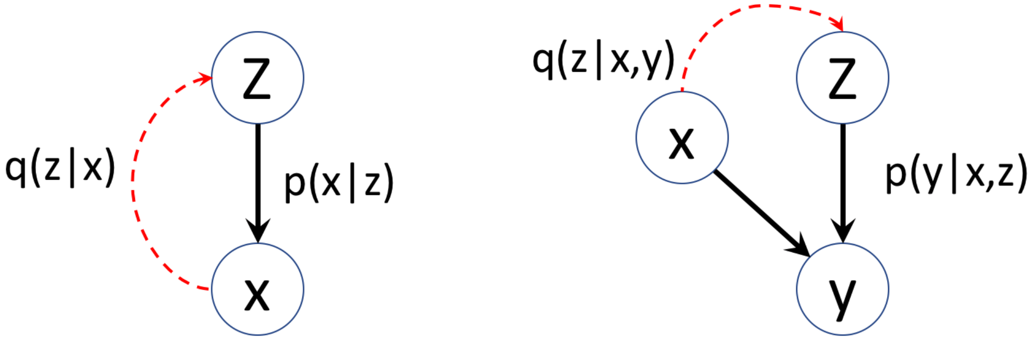


Figure 6.6 ***Graphical probabilistic representation of Variational Autoencoder: left) Variational Autoencoder right) conditional Variational Autoencoder.***

By replacing the $p(z|x)$ from Eq. (6.4) and apply log properties,

$$\begin{aligned}
KL\left(q(z)||p(z|x)\right) &= -\sum q(z)\text{log }\left[\frac{p(x,z)}{q(z)}.\frac{1}{p(x)}\right] \\
&= -\sum q(z)\left[\text{log }\frac{p(x,z)}{q(z)} + \text{log }\frac{1}{p(x)}\right] \\
&= -\sum q(z)\left[\text{log }\frac{p(x,z)}{q(z)} - \text{log }p(x)\right] \qquad (6.6)\\
&= -\sum q(z)\text{log }\frac{p(x,z)}{q(z)} + \sum q(z)\text{log }p(x) \\
&= -\sum q(z)\text{log }\frac{p(x,z)}{q(z)} + \text{log }p(x)\cancelto{1}{\sum q(z)}.
\end{aligned}$$

Let's rewrite the Eq. (6.6), as below:

$$\text{log }p(x) = KL\left(q(z)||p(z|x)\right) + \sum q(z)\text{log }\frac{p(x,z)}{q(z)}, \qquad (6.7)$$

where x is given and known and $p(x)$ is a fixed scalar value. We want to minimize the KL divergence (first term in right-hand). Therefore, we realize that minimizing KL divergence is equal to maximizing the second right-hand term in Eq. (6.7) (we called it $\mathcal{L}$). This term is called variational lower bound. In variational inference we deal with this term and try to maximize it. It's called lower bound because of this fact that KL divergence is always positive, so we can write Eq. (6.7) as below:

$$\mathcal{L} = \sum q(z)\text{log }\frac{p(x,z)}{q(z)} \leq \text{log }p(x), \qquad (6.8)$$

where $\mathcal{L}$ is lower bound of log probability. If we maximize the lower bound, it is like we maximize the log probability. Now, we rewrite variational lower bound as below by replacing it from Eq. (6.4) and apply log properties,

$$\begin{aligned}\mathcal{L} = \sum q(z)\text{log }\frac{p(x,z)}{q(z)} &= \sum q(z)\text{log }\frac{p(x|z)p(z)}{q(z)}\\ &= \sum q(z)\left[\text{log }p(x|z) + \text{log }\frac{p(z)}{q(z)}\right]\\ &= \sum q(z)\text{log }p(x|z) + \sum q(z)\text{log }\frac{p(z)}{q(z)}\\ &= \sum q(z)\text{log }p(x|z) - KL\left(q(z)||p(z)\right).\end{aligned} \tag{6.9}$$

The first term in the right-hand side of Eq. (6.9) is the expectation of log probability of $p(x|z)$ with respect to $q(z)$. Then, we can rewrite Eq. (6.9) as below:

$$\mathcal{L} = \mathbb{E}_{q(z)}\text{log }p(x|z) - KL\left(q(z)||p(z)\right). \tag{6.10}$$

We want to maximize Eq. (6.10) in the variational autoencoder. In fact, we minimize the minus of Eq. (6.10) during the training to optimize weights in the encoder and decoder networks. The loss of conditional variational autoencoder can be calculated by easily extending the Eq. (6.10) to its conditional version (Fig. 6.6) as follow:

$$\mathcal{L} = \mathbb{E}_{q(z|x,y)}\text{log }p(y|z,x) - KL\left(q(z|x,y)||p(z|x)\right). \tag{6.11}$$

When the noise is added to the input ($x- > x'$), we can arrive to the formulation in Eq. (6.3) By combining the Eq. (6.10) and denoising formulations [75] as follow:

$$\mathcal{L}_{CVAE}(\phi,\theta;x',y) = \mathcal{L}_{MSE} + \mathcal{L}_{prior}, \tag{6.12}$$

where

$$\begin{aligned}
&\mathcal{L}_{MSE} = -\mathbb{E}_{p(x'|x)}\mathbb{E}_{q(z|x',y)}[\log p_\theta(y|z,x)], \\
&\mathcal{L}_{prior} = \mathbb{E}_{p(x'|x)}[D_{KL}(\hat{q}_\phi(z|x',y)||p(z|x))], \\
&\hat{q}_\phi(z|x',y) = \mathbb{E}_{p(x'|x)}[q_\phi(z|x',y)].
\end{aligned} \tag{6.13}$$

CHAPTER 7

DISCUSSION AND CONCLUSION

Prior to this thesis, a time-domain optical scanner was developed in our group for small animal optical imaging. The main purpose of this work was to develop practical image reconstruction algorithms to be integrated with this scanner. Two main approaches were considered, iterative model-based and machine learning-based image reconstruction algorithms. In this chapter, gained insights and an appreciation of these approaches will be discussed. Future work to improve the performance of the proposed algorithms will be suggested.

7.1 Thesis review

7.1.1 Iterative model-based algorithms

Diffuse optical tomographic image reconstruction is a difficult problem to solve owing to its ill-posedness and nonlinearity. Iterative model-based image reconstruction methods are one way to solve this problem. In these algorithms a fitting model is defined to find optimal parameters iteratively. These algorithms consist of two problems that need to be solved at each iteration. First, the forward problem must be solved to simulate the physical interaction of light inside the medium with provided optical properties. The output of the forward problem is a set of predicted optical measurements on the surface of the medium. Secondly, the inverse problem needs be solved to update optical properties of medium in order to decrease the discrepancy between the experimental measurements and their prediction by the forward model. For this an objective function is defined, and then an optimization algorithm minimizes the objective function along a descent direction. Depending on the optimization algorithm, the descent direction can be calculated from the gradient/Hessian of the objective function. The output of the inverse problem at a given iteration are the updated optical properties to be used in the forward problem at the next iteration. This process is repeated until the objective function reaches a minimum determined via certain criteria. The final distribution of optical properties is displayed as an image.

Since our optical scanner can acquire time-domain information, it is possible to to use all optical data types (time-domain - TD, frequency-domain - FD, and continuous-wave - CW). In this project, continuous-wave information (sum values of TPSF curves) and full-curve information are used. Two forward models are developed to simulate the light propagation in the medium based on the CW and TD diffusion approximation. In both light propagation models, finite-differences are used to discretize the problem spatially and temporally. For minimizing the objective function, 4 different optimization methods are implemented, one from quasi-Newton optimization methods (L-BFGS algorithm) and 3 from Newton's optimization methods (TNC, LM, and GN algorithms). To investigate the performance of each algorithm, all of them were applied to a CW multispectral DOT problem. The optimization methods were sorted based on their implementation requirements into 2 main groups: gradient-based and Jacobian-based. In the gradient-based group (including L-BFGS and TNC), the gradient of the objective function is calculated with respect to the optical concentrations by extending the objective function to a Lagrange function (adjoint method). In the Jacobian-based group (including LM and GN), the Jacobian matrix of the objective function is calculated with respect to the optical concentrations (reciprocal theory). In the Jacobian-based group, Newton's equation which is very large needs be solved using an iterative solver. A Jacobian-free Newton-Krylov (JFNK) solver is selected. This iterative solver is very efficient for large problems such as the multispectral DOT reconstruction problem.

A full-curve TPSF reconstruction algorithm is developed for image reconstruction from TD measurements. Since each time step in TPSF curves is used in the image reconstruction process, this makes the reconstruction process computationally very demanding. In this work, a combined CW/TD model-based image reconstruction paradigm is proposed for fluorescence lifetime - FLT - tomography, making it less computationally expensive. Since in the reconstruction process the value of the TPSF at each time bin is used, noise can greatly affect the quality of the computation of the gradient. This can lead to failure of the minimization process. To suppress the noise, a filter is applied on the TPSFs to smooth out the fluctuations. Then the smoothed TPSFs are normalized to their energies. For real measurements, this normalization can eliminate experimental factors such as the hard to calibrate optical efficiency of the detection channels. In the reconstruction algorithm developed, first, a CW objective function is defined and minimized with an L-BFGS algorithm to recover the fluorescence yield map. The output of this step is thresholded and fed as a mask to a full-curve image reconstruction algorithm for FLT reconstruction. Using a mask and solving the inverse problem for the nodes inside the mask accelerates the image reconstruction process. Reconstruction of fluorescence yield with the

CW information and FLT with TD information makes the whole process much shorter, while the whole information of the TPSF curves is extracted and efficiently exploited in regions of interest.

To validate the FLT image reconstruction algorithm, two numerical experiments were performed. Since full-curve information is used in the reconstruction process, it has the potential to reconstruct several FLTs. The results show that 4 FLT inclusions with different FLTs can be localized and quantified even when the distance between them is small. To evaluate the algorithm for a realistic or complex geometry, a numerical mouse phantom is used. Three different lifetimes were seated in the heart, lung, and liver of the mouse phantom. The reconstructed images displayed sharp edges, allowing to distinguish different organs with high contrast. The whole reconstruction time was about 8 hours for the mouse phantom, which is a reasonable time for time-domain small animal image reconstruction.

As a conclusion, some of the advantages and limitations of iterative model-based algorithms for (F)DOT are the following:

- Advantages:
  - A general technique which allows reconstructing any arbitrary medium based on solid physical and mathematical principles.
- Limitations:
  - Leads to time consuming algorithms.
  - Requires precise geometrical information about the medium.
  - Requires providing an initial map to start the optimization algorithm.

7.1.2 Machine learning-based algorithm

An (F)DOT image reconstruction algorithm based on machine learning techniques was proposed as an alternative to iterative model-based algorithms. Among all machine learning techniques, deep learning approaches have drawn a lot of attention in the last few years with their ability to efficiently solve highly complex problems. Deep learning models are able to model highly nonlinear processes through deeper architectures. An interesting model in the deep learning for image reconstruction is the generative model, which is applied successfully for image generation in several applications. In this work, the performance of deep generative models for FDOT image reconstruction has been evaluated. The deep generative model can learn an optimal inversion function between measurement and image space. The proposed model is adapted from the variational autoencoder -

VAE - model which is a well-known model for unsupervised learning and variational inference. VAEs consist of encoder and decoder networks. The encoder network maps the measurements from a high-dimensional space to a low-dimensional joint-manifold space. By sampling from that space, the decoder can reconstruct an image. The model is first trained with a bulk of data which was here synthetically generated from a forward model (note that real data could also be used for that purpose). After training, it can reconstruct a slice image in less than 1 second. In the present case, the training time was around 8 hours. Hence, once trained, the proposed model shows great potential and paves the way for real-time diffuse optical tomographie imaging applications, which are not yet possible.

The proposed algorithm showed great robustness to noise and sparse measurements, which is also not possible with current image reconstruction techniques. First, the model was trained with synthetic data, and then was tested with experimental data. The recovered image features were well localized even for very sparse measurements. In the second experiment, the model's performance was evaluated for a more general problem. In this case, the model was trained for two different cylindrical volumes with different sizes. Then it was asked to reconstruct a volume with an unknown size between the two training volumes. This shows the capability of the model to generalize to objects of sizes other than those used for training. Such a capability is not possible with current standard image reconstruction techniques, whereby the geometry must be beforehand to the reconstruction algorithm.

Deep learning models, such as the model proposed here, can be potentially extended to other imaging modalities since the learned function is independent of the data acquisition strategy. One limitation for learning algorithms is preparing a big dataset for each medium with different sizes and shapes. In the medical imaging field, this limitation can be compensated by developing precise forward models to simulate measurements made on the medium during scanning to generate synthetic data for training deep learning models.

As a conclusion, some of the advantages and limitations of machine learning techniques for (F)DOT image reconstruction are the following:

- Advantages:
  - Fast reconstructing of a whole 3D volume (suitable for real-time tomography).
  - High robustness against noisy measurement.
  - Ability to deal with sparse measurements.
  - Geometry information is not a necessity here.

- Limitations:
  - Requires lots of data and measurements for training.
  - Requires a broad range of data. A low-biased and high-variance dataset should be generated synthetically to improve the model generalization.

7.2 Thesis contributions

7.2.1 Iterative model-based algorithms

A CW model-based image reconstruction algorithm was developed for 3D multispectral DOT imaging. Several optimization algorithms, including Newton and quasi-Newton methods, were implemented. These algorithms were evaluated and compared for several conditions that differ in terms of the scale of the problem, the starting point (initial guess) and noise level. The performance of each implemented optimization algorithm was evaluated for a large problem such as a CW multispectral DOT problem. To the best of the author's knowledge, this is the first time such a vast range of techniques are applied to a CW multispectral DOT problem.

A TD FLT image reconstruction was developed for localizing and quantifying several different FLTs simultaneously, allowing to discriminate between fluorophores that may have similar emission spectra, but different FLTs. This is a main reason why FLT tomography is of high interest. Full-curve TPSF information has never been used for a large 3D big volume such as a mouse for FLT imaging because of the computational costs imposed by handling whole TPSF curves during iterative image reconstruction. The proposed algorithm here accelerates the reconstruction process to a reasonable time for reconstructing a mouse-size phantom. This makes the algorithm suitable for practical applications.

7.2.2 Machine learning-based algorithms

A novel machine learning-based image reconstruction algorithm was proposed for fluorescence optical tomography. It was the first time that these algorithms were applied in FDOT. Moreover, the algorithm developed here could easily be modified and adapted to other imaging modalities, such as nuclear medicine, since a low-dimensional joint distribution between the sensor domain and image domain can be learned, independent from the data acquiring strategy.

7.3 Future work

Several suggestions are proposed here to improve the performance of (F)DOT image reconstruction algorithms. Firstly, some suggestions are proposed to improve techniques used in (F)DOT image reconstruction algorithms. Secondly, suggestions are provided to improve the implementation of (F)DOT image reconstruction algorithms.

Technique

— Using multi-grid meshing techniques in the CW forward model to improve the precision of predicted light fields on the boundary of the medium, while keeping the computation time intact.

— Accelerating TD forward model by replacing Crank-Nicolson methods with other methods such as alternating direction implicit (ADI) method.

— Modifying the objective function from an L_2 to another form such as L_1, L_p $(0 < p < 1)$ for FLT to enhance image sparsity.

— Modifying the loss function of deep generative model to enhance the quality of reconstructed image.

— Applying deep generative model in other areas of optical imaging such as intrinsic, multispectral, and bioluminescence imaging,

— Modifying the architecture of the deep generative model for other type of optical measurements such as time-domain and frequency-domain,

— Extending 2D-CNN architecture to a 3D-CNN to generate whole 3D image in one stage.

Implementation

— Using CUDA libraries for linear algebra computation to accelerate the processing time.

— Extend the deep generative model to a distributed platform to accelerate the training time.

— Extend the deep generative model to a distributed platform to be able to implement 3D-CNN architectures.

LIST OF REFERENCES

[1] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M. *et al.* (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

[2] Abascal, J. F. P., Aguirre, J., Chamorro-Servent, J., Schweiger, M., Arridge, S. R., Ripoll, J., Vaquero, J. J. and Desco, M. (2012). Influence of absorption and scattering on the quantification of fluorescence diffuse optical tomography using normalized data. *Journal of Biomedical Optics*, volume 17(3), pp. 036013–(1–9).

[3] Achilefu, S. (2004). Lighting up tumors with receptor-specific optical molecular probes. *Technol. Cancer Res. Treat.*, volume 3, pp. 393–409.

[4] Alander, J. T., Kaartinen, I., Laakso, A., Pätilä, T., Spillmann, T., Tuchin, V. V., Venermo, M. and Välisuo, P. (2012). A review of indocyanine green fluorescent imaging in surgery. *Journal of Biomedical Imaging*, volume 2012, p. 7.

[5] Antholzer, S., Haltmeier, M. and Schwab, J. (2017). Deep learning for photoacoustic tomography from sparse data. *arXiv preprint arXiv:1704.04587*.

[6] Arridge, S. (1999). Optical tomography in medical imaging. *Inverse Problems*, volume 15, number 2, pp. 41–93.

[7] Arridge, S. R. and Hebden, J. C. (1997). Optical imaging in medicine: II. Modelling and reconstruction. *Physics in Medicine and Biology*, volume 42, number 5, p. 841.

[8] Arridge, S. R. and Lionheart, W. R. B. (1998). Nonuniqueness in diffusion-based optical tomography. *Opt. Lett.*, volume 23, number 11, pp. 882–884.

[9] Arridge, S. R. and Schotland, J. C. (2009). Optical tomography: Forward and inverse problems. *Inverse Problems*, volume 25, number 12, pp. 123–133.

[10] Arridge, S. R. and Schweiger, M. (1995). Photon-measurement density functions. Part 2: Finite-element-method calculations. *Appl. Opt.*, volume 34, number 34, pp. 8026–8037.

[11] Arridge, S. R. and Schweiger, M. (1997). Image reconstruction in optical tomography. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, volume 352, number 1354, pp. 717–726.

[12] Arridge, S. R. and Schweiger, M. (1998). A gradient-based optimisation scheme for optical tomography. *Optics Express*, volume 2, pp. 213–226.

[13] Arridge, S. R., van der Zee, P., Cope, M. and Delpy, D. T. (1991). Reconstruction methods for infrared absorption imaging. *Proc. SPIE*, volume 1431, pp. 204–215.

[14] Austin, T., Gibson, A., Branco, G., Yusof, R., Arridge, S., Meek, J., Wyatt, J., Delpy, D. and Hebden, J. (2006). Three dimensional optical imaging of blood volume and oxygenation in the neonatal brain. *Neuroimage*, volume 31, pp. 1426–1433.

[15] Baker, A. H., Jessup, E. R. and Manteuffel, T. (2005). A technique for accelerating the convergence of restarted GMRES. *SIAM Journal on Matrix Analysis and Applications*, volume 26, number 4, pp. 962–984.

[16] Bar, Y., Diamant, I., Wolf, L. and Greenspan, H. (2015). Deep learning with non-medical training used for chest pathology identification. In *Proc. SPIE*. volume 9414. p. 94140V.

[17] Becker, W. (2010 (4th Ed.)). *The bh TCSPC Handbook*. Becker&Hickl GmbH, Berlin, Germany.

[18] Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, volume 2, number 1, pp. 1–127.

[19] Berezin, M. Y. and Achilefu, S. (2010). Fluorescence lifetime measurements and biological imaging. *Chemical Reviews*, volume 110, number 5, pp. 2641–2684.

[20] Bérubé-Lauzière, Y. (2015). Diffuse optical tomography course notes.

[21] Boas, D. (1997). A fundamental limitation of linearized algorithms for diffuse optical tomography. *Optics Express*, volume 1, number 13, pp. 404–413.

[22] Boas, D., Brooks, D., Miller, E., DiMarzio, C., Kilmer, M., Gaudette, R. and Zhang, Q. (2001). Imaging the body with diffuse optical tomography. *IEEE Signal Processing Magazine*, volume 18, number 6, pp. 57–75.

[23] Boas, D., Liu, H., O'Leary, M., Chance, B. and Yodh, A. G. (1995). Photon migration within the SP3 approximation. *Proc. SPIE*, volume 2389, pp. 240–246.

[24] Bouza-Domínguez, J. and Bérubé-Lauzière, Y. (2011). Light propagation from fluorescent probes in biological tissues by coupled time-dependent parabolic simplified spherical harmonics equations. *Biomedical Optics Express*, volume 2, number 4, pp. 817–837.

[25] Bouza Domínguez, J. and Bérubé-Lauzière, Y. (2012). Diffuse optical tomographic imaging of biological media by time-dependent parabolic SP_N equations: a two-dimensional study. *Journal of Biomedical Optics*, volume 17, number 8, pp. 086012–1–14.

[26] Bremer, C., Ntziachristos, V. and Weissleder, R. (2003). Optical-based molecular imaging:contrast agents and potential medical applications. *European Radiology*, volume 13, number 2, pp. 231–243.

[27] Brosch, T., Tang, L. Y., Yoo, Y., Li, D. K., Traboulsee, A. and Tam, R. (2016). Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Transactions on Medical Imaging*, volume 35, number 5, pp. 1229–1239.

[28] Chavent, G. (2010). *Nonlinear Least Squares for Inverse Problems*. Springer.

[29] Chen, J., Venugopal, V. and Intes, X. (2011). Monte Carlo based method for fluorescence tomographic imaging with lifetime multiplexing using time gates. *Biomed. Opt. Express*, volume 2, number 4, pp. 871–886.

[30] Cheong, W., Prahl, S. A. and Welch, A. J. (1990). A review of the optical properties of biological tissues. *IEEE Journal of Quantum Electronics*, volume 26, number 12, pp. 2166 – 2185.

[31] Ciarlet, P. G. and Lions, J. L. (2003). *Handbook of Numerical Analysis*. Elsevier Science B. V.

[32] Corlu, A., Choe, R., Durduran, T., Lee, K., Schweiger, M., Arridge, S. R., Hillman, E. M. and Yodh, A. G. (2005). Diffuse optical tomography with spectral constraints and wavelength optimization. *Applied optics*, volume 44, number 11, pp. 2082–2093.

[33] Corlu, A., Choe, R., Durduran, T., Rosen, M., Schweiger, M., Arridge, S., Schnall, M. and Yodh, A. (2007). Three dimensional in vivo fluorescence diffuse optical tomography of breast cancer in humans. *Optics Express*, volume 15, pp. 6696–6716.

[34] Corlu, A., Durduran, T., Choe, R., Schweiger, M., Hillman, E. M. C., Arridge, S. R. and Yodh, A. G. (2003). Uniqueness and wavelength optimization in continuous-wave multispectral diffuse optical tomography. *Opt. Lett.*, volume 28, number 23, pp. 2339–2341.

[35] Correia, T., Aguirre, J., Sisniega, A., Chamorro-Servent, J., Abascal, J., Vaquero, J. J., Desco, M., Kolehmainen, V. and Arridge, S. (2011). Split operator method for fluorescence diffuse optical tomography using anisotropic diffusion regularisation with prior anatomical information. *Biomedical Optics Express*, volume 2(9), pp. 2632–2648.

[36] Cristea, P. D. (2009). Application of neural networks in image processing and visualization. In *GeoSpatial Visual Analytics*. Springer, pp. 59–71.

[37] Cubeddu, R., Comelli, D., D'Andrea, C., Taroni, P. and Valentini, G. (2002). Time-resolved fluorescence imaging in biology and medicine. *Applied Optics*, volume 35, pp. 61–76.

[38] Culver, J. P., Durduran, T., Cheung, C., Yodh, A. G., Furuya, D. and Greenberg, J. H. (2003). *Diffuse Optical Measurement of Hemoglobin and Cerebral Blood Flow in Rat Brain During Hypercapnia, Hypoxia and Cardiac Arrest*. Springer US, Boston, MA, pp. 293–297.

[39] Darne, C., Lu, Y. and Sevick-Muraca, E. M. (2014). Small animal fluorescence and bioluminescence tomography: A review of approaches, algorithms and technology update. *Physics in Medicine and Biology*, volume 59, number 1, p. R1.

[40] Dehghani, H., Eames, M. E., Yalavarthy, P. K., Davis, S. C., Srinivasan, S., Carpenter, C. M., Pogue, B. W. and Paulsen, K. D. (2009). Near infrared optical tomography using NIRFAST: Algorithm for numerical model and image reconstruc-

tion. *Communications in Numerical Methods in Engineering*, volume 25, number 6, pp. 711–732.

[41] Dehghani, H., Srinivasan, S., Pogue, B. W. and Gibson, A. (2009). Numerical modelling and image reconstruction in diffuse optical tomography. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 367, number 1900, pp. 3073–3093.

[42] Diagaradjane, P., Orenstein-Cardona, J. M., E. Colèn-Casasnovas, N., Deorukhkar, A., Shentu, S., Kuno, N., Schwartz, D. L., Gelovani, J. G. and Krishnan, S. (2008). Imaging epidermal growth factor receptor expression in vivo: Pharmacokinetic and biodistribution characterization of a bioconjugated quantum dot nanoprobe. *Clinical Cancer Research*, volume 14, number 3, pp. 731–741.

[43] Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.

[44] Dogdas, B., Stout, D., Chatziioannou, A. F. and Leahy, R. M. (2007). Digimouse: A 3D whole body mouse atlas from CT and cryosection data. *Physics in Medicine and Biology*, volume 52, number 3, p. 577.

[45] Domínguez, J. B. and Bérubé-Lauzière, Y. (2010). Diffuse light propagation in biological media by a time-domain parabolic simplified spherical harmonics approximation with ray-divergence effects. *Applied optics*, volume 49, number 8, pp. 1414–1429.

[46] Douiri, A., Schweiger, M., Riley, J. and Arridge, S. R. (2007). Anisotropic diffusion regularization methods for diffuse optical tomography using edge prior information. *Measurement Science and Technology*, volume 18, pp. 87–95.

[47] Dutta, J., Ahn, S., Li, C., Cherry, S. R. and Leahy, R. M. (2012). Joint L1 and total variation regularization for fluorescence molecular tomography. *Physics in Medicine and Biology*, volume 57, number 6, p. 1459.

[48] Edjlali, E. and Bérubé-Lauzière, Y. (2016). Analytical solution of the simplified spherical harmonics equations in spherical turbid media. *Journal of Quantitative Spectroscopy and Radiative Transfer*, volume 182, pp. 112–118.

[49] E.Nishani and Cico, B. (2017). Computer vision approaches based on deep learning and neural networks: Deep neural networks for video analysis of human pose estimation. In *2017 6th Mediterranean Conference on Embedded Computing (MECO)*. pp. 1–4.

[50] Ettehadi, S. and Bérubé-Lauziére, Y. (2017). Enhancing fluorescence lifetime image reconstruction through TPSF full-curve information. In *Diffuse Optical Spectroscopy and Imaging VI*. Optical Society of America, p. 104120X.

[51] Fessler, J. A. and Sutton, B. P. (2003). Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing*, volume 51, number 2, pp. 560–574.

[52] Fletcher, R. (1987). *Practical Methods of Optimization; (2nd Ed.).* Wiley-Interscience, New York, NY, USA.

[53] Freyer, M., Ale, A., Schulz, R. B., Zientkowska, M., Ntziachristos, V. and Englmeier, K. (2010). Fast automatic segmentation of anatomical structures in x-ray computed tomography images to improve fluorescence molecular tomography reconstruction. *Journal of Biomedical Optics*, volume 15(3), p. 036006.

[54] Gao, F., Li, J., Zhang, L., Poulet, P., Zhao, H. and Yamada, Y. (2010). Simultaneous fluorescence yield and lifetime tomography from time-resolved transmittances of small-animal-sized phantom. *Applied Optics*, volume 49, number 16, pp. 3163–3172.

[55] Gao, F., Zhao, H., Tanikawa, Y. and Yamada, Y. (2006). A linear, featured-data scheme for image reconstruction in time-domain fluorescence molecular tomography. *Optics Express*, volume 14(16), pp. 7109–7124.

[56] Ghesu, F. C., Krubasik, E., Georgescu, B., Singh, V., Zheng, Y., Hornegger, J. and Comaniciu, D. (2016). Marginal space deep learning: Efficient architecture for volumetric image parsing. *IEEE Transactions on Medical Imaging*, volume 35, number 5, pp. 1217–1228.

[57] Gibson, A. and Dehghani, H. (2009). Diffuse optical imaging. *Phil. Trans. R. Soc. A*, pp. 3055–3072.

[58] Gibson, A., Hebden, J. and Arridge, S. (2005). Recent advances in diffuse optical imaging. *Phys. Med. Biol.*, volume 50, pp. 1–43.

[59] Godavarty, A., Sevick-Muraca, E. M. and Eppstein, M. J. (2005). Three-dimensional fluorescence lifetime tomography. *Medical Physics*, volume 32, number 4, pp. 992–1000.

[60] Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*. MIT Press.

[61] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*. pp. 2672–2680.

[62] Gulsen, G., Xiong, B., Birgul, O. and Nalcioglu, O. (2006). Design and implementation of a multifrequency near-infrared diffuse optical tomography system. *Journal of Biomedical Optics*, volume 11.

[63] Haller, J., Hyde, D., Deliolanis, N., de Kleine, R., Niedre, M. and Ntziachristos, V. (2008). Visualization of pulmonary inflammation using noninvasive fluorescence molecular imaging. *J. Appl. Physiol.*, volume 104, pp. 795–802.

[64] Haskell, R. C., Svaasand, L. O., Tsay, T.-T., Feng, T.-C., Tromberg, B. J. and McAdams, M. S. (1994). Boundary conditions for the diffusion equation in radiative transfer. *J. Opt. Soc. Am. A*, volume 11, number 10, pp. 2727–2741.

[65] Hawrysz, D. and Sevick-Muraca, E. (2000). Developments toward diagnostic breast cancer imaging using near-infrared optical measurements and fluorescent contrast agents. *Neoplasia*, volume 2, number 5, pp. 388–417.

[66] Hebden, J., Yates, T., Gibson, A., Everdel, l. N., Arridge, S., Chicken, D., Douek, M. and Keshtgar, M. . (2005). Monitoring recovery after laser surgery of the breast with optical tomography: a case study. *Appl. Opt.*, volume 44, pp. 1898–1904.

[67] Hielscher, A. (2005). Optical tomographic imaging of small animals. *Current Opinion in Biotechnology*, volume 16, number 1, pp. 79–88.

[68] Hielscher, A., Klose, A. D.and Scheel, A., Moa-Anderson, B., Backhaus, M., Netz, U. and Beuthan, J. (2004). Sagittal laser optical tomography for imaging of rheumatoid finger joints. *Physics in Medicine and Biology*, volume 49, p. 1147.

[69] Hielscher, A. H., Alcouffe, R. E. and Barbour, R. (1998). Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues. *Physics in Medicine and Biology*, volume 43, pp. 1285–1302.

[70] Hielscher, A. H. and Bartel, S. (2001). Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography. *J. Biomed. Opt.*, volume 6, pp. 183–192.

[71] Hielscher, A. H., Klose, A. D. and Hanson, K. M. (1999). Gradient-Based Iterative Image Reconstruction Scheme for Time-Resolved Optical Tomography. *IEEE Transactions on Medical Imaging*, volume 18(3), pp. 262–271.

[72] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N. *et al.* (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, volume 29, number 6, pp. 82–97.

[73] Hou, S., Rice, W. L., Bacskai, B. and Kumar, A. T. (2014). Hybrid reconstruction algorithms for time domain diffuse fluorescence tomography. In *Biomedical Optics 2014*. Optical Society of America, p. BM3A.48.

[74] Housley, J. (2010). *Determination of the Optimum Metrics for the Characterisation of Scattering Media*. Ph.D. thesis, University of Nottingham.

[75] Im, D. J., Ahn, S., Memisevic, R., Bengio, Y. *et al.* (2017). Denoising criterion for variational auto-encoding framework. In *AAAI*. pp. 2059–2065.

[76] Izmailova, E., Paz, N., Alencar, H., Chun, M., L., S., Hepperle, M., Lane, J., Harriman, G. X., Y, Ocain, T., Weissleder, R., Mahmood, U., Healy, A. and Jaffee, B. (2007). Use of molecular imaging to quantify response to ikk-2 inhibitor treatment in murine arthritis. *Arthritis Rheum*, volume 56, pp. 117–128.

[77] Jares-Erijman, E. A. and Jovin, T. M. (2006). Imaging molecular interactions in living cells by FRET microscopy. *Current Opinion in Chemical Biology*, volume 10, number 5, pp. 409–416.

[78] Jiang, H., Paulsen, K. D., Osterberg, U. L., Pogue, B. W. and Patterson, M. S. (1996). Optical image reconstruction using frequency-domain data: simulations and experiments. *J. Opt. Soc. Am. A*, volume 13, pp. 253–266.

[79] Jiang, H. B. (1999). Optical image reconstruction based on the third-order diffusion equations. *Opt. Express*, volume 4, pp. 241–246.

[80] Jobsis, F. F. (1977). Noninvasive infrared monitoring of cerebral and myocardial sufficiency and circulatory parameters. *Science*, volume 198, pp. 1264–1267.

[81] Jones, E., Oliphant, T., Peterson, P. *et al.* (2001–). SciPy: Open source scientific tools for Python.

[82] Kak, A. C. and Slaney, M. (2001). *Principles of computerized tomographic imaging.* SIAM.

[83] Kang, E., Min, J. and Ye, J. C. (2017). A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Medical Physics*, volume 44, number 10.

[84] Karpathy, A. (2017). CS231n: Convolutional neural networks for visual recognition. `https://github.com/cs231n/cs231n.github.io`.

[85] Kim, H. and Charette, A. (2007). A sensitivity function-based conjugate gradient method for optical tomography with the frequency-domain equation of radiative transfer. *Journal of Quantitative Spectroscopy and Radiative Transfer*, volume 104, pp. 24–39.

[86] Kinahan, P. E. and Fletcher, J. W. (2010). Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. In *Seminars in Ultrasound, CT and MRI*, Elsevier. volume 31. pp. 496–505.

[87] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.*

[88] Kingma, D. P. and Welling, M. (2013). Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114.*

[89] Klose, A. and Hielscher, A. (2002). Optical tomography using the time-independent equation of radiative transfer - part 1: Forward model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, volume 72, number 5, pp. 691–713.

[90] Klose, A. and Hielscher, A. (2002). Optical tomography using the time-independent equation of radiative transfer - part 2: Inverse model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, volume 72, number 5, pp. 715 – 32.

[91] Klose, A. and Hielscher, A. H. (2006). Fluorescence molecular tomography of small animals using the radiative transfer equation for curved geometries. *Proceedings of the OSA Biomedical Topical Meeting, Fort Lauderdale, FL.*

[92] Klose, A., Moa-Anderson, B. and Hielscher, A. (2003). Near-infrared fluorescence molecular tomography. *Optics Letters*, volume 28, pp. 1019–1021.

[93] Klose, A. D., Beattie, B. J., Dehghani, H., Vider, L., Le, C., Ponomarev, V. and Blasberg, R. (2010). In vivo bioluminescence tomography with a blocking-off finite-difference SP3 method and MRI/CT coregistration. *Medical Physics*, volume 37(1), pp. 329–338.

[94] Klose, A. D. and Hielscher, A. H. (1999). Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer. *The Journal of The American Association of Physicists in Medicine*, volume 26(8), pp. 1698–1707.

[95] Klose, A. D. and Hielscher, A. H. (2003). Quasi-Newton methods in optical tomographic image reconstruction. *Inverse Problems*, volume 19, number 2, p. 387.

[96] Klose, A. D., Hielscher, A. H. and Beuthan, J. (2000). Quasi-Newton methods in iterative image reconstruction schemes for optical tomography. *Proc. SPIE*, volume 3979, pp. 1323–1330.

[97] Klose, A. D. and Larsen, E. W. (2006). Light transport in biological tissue based on the simplified spherical harmonics equations. *J. Comput. Phys.*, volume 220, pp. 441–470.

[98] Knoll, D. and Keyes, D. (2004). Jacobian-free Newton-Krylov methods: a survey of approaches and applications. *Journal of Computational Physics*, volume 193, number 2, pp. 357 – 397.

[99] Kobayashi, H., Hama, Y., Koyama, Y., Barrett, T., Regino, C. A., Urano, Y. and Choyke, P. L. (2007). Simultaneous multicolor imaging of five different lymphatic basins using quantum dots. *Nano Letters*, volume 7, number 6, pp. 1711–1716.

[100] Kosaka, N., Ogawa, M., Sato, N., Choyke, P. L. and Kobayashi, H. (2009). In vivo real-time, multicolor, quantum dot lymphatic imaging. *Journal of Investigative Dermatology*, volume 129, number 12, pp. 2818–2822.

[101] Kozlo, K. M., Weissleder, R. and Mahmood, U. (2008). Noninvasive optical detection of bone mineral. *Journal of Bone and Mineral Research*, volume 22, number 8, pp. 1208–1216.

[102] Kulkarni, T. D., Whitney, W., Kohli, P. and Tenenbaum, J. B. (2015). Deep convolutional inverse graphics network. *CoRR*, volume abs/1503.03167.

[103] Kumar, A., Raymond, S., Boverman, G., Boas, D. and Bacskai, B. (2006). Time resolved fluorescence tomography of turbid media based on lifetime contrast. *Optics Express*, volume 14, number 25, pp. 12255–12270.

[104] Kumar, A. T., Carp, S., Yang, J., Ross, A., Medarova, Z. and Ran, C. (2017). Fluorescence lifetime-based contrast enhancement of indocyanine green-labeled tumors. *Journal of Biomedical Optics*, volume 22, number 4, p. 040501.

[105] Kumar, A. T., Raymond, S. B., Bacskai, B. J. and Boas, D. A. (2008). Comparison of frequency-domain and time-domain fluorescence lifetime tomography. *Optics Letters*, volume 33, number 5, pp. 470–472.

[106] Kumar, A. T., Skoch, J., Bacskai, B. J., Boas, D. A. and Dunn, A. K. (2005). Fluorescence-lifetime-based tomography for turbid media. *Optics Letters*, volume 30, number 24, pp. 3347–3349.

[107] Kumar-Anand, T., Raymond, S. B., Dunn, A. K., Bacskai, B. J. and Boas, D. A. (2008). A time domain fluorescence tomography system for small animal imaging. *IEEE Transactions on Medical Imaging*, volume 27, number 8, pp. 1152–1163.

[108] Lakowicz, J. R. (2006). *Principles of Fluorescence Spectroscopy*, 2nd edition. Springer.

[109] Lam, S., Lesage, F. and Intes, X. (2005). Time domain fluorescent diffuse optical tomography: Analytical expressions. *Optics Express*, volume 13, number 7, pp. 2263–2275.

[110] Lapointe, E., Pichette, J. and Bérubé-Lauzière, Y. (2012). A multi-view time-domain non-contact diffuse optical tomography scanner with dual wavelength detection for intrinsic and fluorescence small animal imaging. *Review of Scientific Instruments*, volume 83, number 6, 063703.

[111] Larsen, A. B. L., Sønderby, S. K., Larochelle, H. and Winther, O. (2015). Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300.*

[112] Larusson, F., Fantini, S. and Miller, E. L. (2011). Hyperspectral image reconstruction for diffuse optical tomography. *Biomedical Optics Express*, volume 2(4), pp. 946–965.

[113] Lebel, R. and Lepage, M. (2014). A comprehensive review on controls in molecular imaging: Lessons from MMP-2 imaging. *Contrast Media & Molecular Imaging*, volume 9, number 3, pp. 187–210.

[114] Leblond, F., Davis, S., Valdès, P. and Pogue, B. (2010). Pre-clinical whole-body fluorescence imaging: Review of instruments, methods and applications. *Journal of Photochemistry and Photobiology B: Biology*, volume 98, pp. 77–94.

[115] Li, X., O'Leary, M., Boas, D. and Chance, B. (1996). Fluorescent diffuse photon density waves in homogeneous and heterogeneous turbid media: Analytic solutions and applications. *Applied Optics*, volume 35, number 19, pp. 3746 – 58.

[116] Licha, K. and Olbrich, C. (2005). Optical imaging in drug discovery and diagnostic applications. *Adv. Drug. Deliv.*, volume 57, pp. 1087–1108.

[117] Mahmood, U. (2004). Near-infrared optical applications in molecular imaging. *IEEE Engineering in Medicine and Biology Magazine*, pp. 58–66.

[118] Montcel, B. and Poulet, P. (2006). An instrument for small-animal imaging using time-resolved diffuse and fluorescence optical methods. *Nuclear Instruments and Methods in Physics Research A*, volume 569, pp. 551–566.

[119] Montejo, L. D., Kima, K. K. and Hielscher, A. H. (2011). A Finite-Volume Algorithm for Modeling Light Transport with the Time-Independent Simplified Spherical

Harmonics Approximation to the Equation of Radiative Transfer. *Proceedings of the international society for optics and photonics (SPIE)*, volume 7896, pp. 78960J–(1–6).

[120] Moscoso, M. *et al.* (2009). Fluorescence lifetime imaging from time resolved measurements using a shape-based approach. *Optics Express*, volume 17, number 11, pp. 8843–8855.

[121] Müller, G. J., Chance, B., Alfano, R. R., Arridge, S. R., Beuthan, J., Gratton, E., Kaschke, M., Masters, B., Svanberg, S. and van der Zee, P. (1993). *Medical Optical Tomography: Functional Imaging and Monitoring*, SPIE institutes for advanced optical technologies Institute Series, *volume 11*. SPIE Institutes for Advanced Optical Technologies, SPIE Optical Engineering Press.

[122] Niedre, M., de Kleine, R., Aikawa, E., Weissleder, R. and Ntziachristos, V. (2008). Early photon tomography allows detection of lung carcinomas and disease progression in mice in vivo. *National Academy of Sciences USA*, volume 105, pp. 19126–19131.

[123] Nocedal, J. and Wright, S. (2006). *Numerical Optimization.* Springer.

[124] Nothdurft, R. E., Patwardhan, S. V., Akers, W., Ye, Y., Achilefu, S. and Culver, J. P. (2009). In vivo fluorescence lifetime tomography. *Journal of Biomedical Optics*, volume 14, number 2, pp. 024–028.

[125] Ntziachristos, V. (2006). Fluorescence molecular imaging. *Annual Review of Biomedical Engineering*, volume 8, pp. 1–33.

[126] Ntziachristos, V., Bremer, C., Graves, E., Ripoll, J. and Weissleder, R. (2002). In vivo tomographic imaging of near-infrared fluorescent probes. *Molecular Imaging*, volume 1, number 2, pp. 82–89.

[127] O'Leary, M. A., Boas, D. A., Chance, B. and Yodh, A. G. (1995). Experimental images of heterogeneous turbid media by frequency-domaindiffusing-photon tomography. *Opt. Lett.*, volume 20, number 5, pp. 426–428.

[128] O'Leary, M. A., Boas, D. A., Li, X. D., Chance, B. and Yodh, A. G. (1996). Fluorescence lifetime imaging in turbid media. *Opt. Lett.*, volume 21, number 2, pp. 158–160.

[129] Paithankar, D. Y., Chen, A. U., Pogue, B. W., Patterson, M. S. and Sevick-Muraca, E. M. (1997). Imaging of fluorescent yield and lifetime from multiply scattered light reemitted from random media. *Appl. Opt.*, volume 36, number 10, pp. 2260–2272.

[130] Pereira, S., Pinto, A., Alves, V. and Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, volume 35, number 5, pp. 1240–1251.

[131] Pichette, J., Goyette, A., Soulez, G., Wilson, B. C. and Leblond, F. (2015). Development of a multispectral Monte Carlo simulation technique for blood vessels detection during brain needle biopsy procedures. In *Optics and the Brain*, Optical Society of America. pp. BrW1B–6.

[132] Prahl, S. A. (2001). Optical Properties Spectra. http://omlc.ogi.edu/spectra.

[133] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*, 3rd edition. Cambridge University Press.

[134] Profio, A. (1989). Light transport in tissue. *Applied Optics*, volume 28, number 12, pp. 2216–2222.

[135] Rafael Gonzalez, C., Woods, E. R. and Eddins, L. S. (2003). Digital image processing using MATLAB. *Prentice Hall, USA., ISBN*, volume 10, p. 0130085197.

[136] Reed, S. E., Akata, Z., Yan, X., Logeswaran, L., Schiele, B. and Lee, H. (2016). Generative adversarial text to image synthesis. *CoRR*, volume abs/1605.05396.

[137] Roy, R. and Sevick-Muraca, E. M. (1999). Truncated Newton's optimization scheme for absorption and fluorescence optical tomography: Part I theory and formulation. *Opt. Express*, volume 4, number 10, pp. 353–371.

[138] Rudin, L. I., Osher, S. and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, volume 60, number 1, pp. 259 – 268.

[139] Sadiku, M. N. O. (2000). *Numerical Techniques in Electromagnetics*. CRC Press LLC.

[140] Sahiner, B., Chan, H.-P., Petrick, N., Wei, D., Helvie, M. A., Adler, D. D. and Goodsitt, M. M. (1996). Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images. *IEEE Transactions on Medical Imaging*, volume 15, number 5, pp. 598–610.

[141] Salo, D. (2017). TensorFlow-VAE. `https://github.com/dancsalo/TensorFlow-VAE`.

[142] Savitzky, A. and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, volume 36, number 8, pp. 1627–1639.

[143] Schmidhuber, J. (2014). Deep learning in neural networks: An overview. *CoRR*, volume abs/1404.7828.

[144] Schulz, R., Ripoll, J., Yessayan, D. and Ntziachristos, V. (2004). Non-contact fluorescence molecular tomography (FMT) of small animals. *Optical Society of America*, volume Paper WA3.

[145] Schweiger, M., Arridge, S. R. and Nissila, I. (2005). Gauss-Newton method for image reconstruction in diffuse optical tomography. *Physics in Medicine and Biology*, volume 50, pp. 2365–2386.

[146] Selb, J., Dale, A. M. and Boas, D. A. (2007). Linear 3D reconstruction of time-domain diffuse optical tomography imaging differential data: Improved depth localization and lateral resolution. *Optics Express*, volume 15(25), pp. 16400–16412.

[147] Setio, A. A. A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S. J., Wille, M. M. W., Naqibullah, M., Sánchez, C. I. and van Ginneken, B. (2016). Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, volume 35, number 5, pp. 1160–1169.

[148] Sevick-Muraca, E. (2004). Fluorescence-enhanced optical imaging and tomography for cancer diagnostics. *IEEE Symposium on Biomedical Imaging*, volume 2, pp. 1482–1487.

[149] Sirinukunwattana, K., Raza, S. E. A., Tsang, Y.-W., Snead, D. R., Cree, I. A. and Rajpoot, N. M. (2016). Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging*, volume 35, number 5, pp. 1196–1206.

[150] Snoeks, T., Lawik, C. and Kaijzel, E. (2010). 'In vivo' optical approaches to angiogenesis imaging. *Angiogenesis*, volume 13, number 2, pp. 135–147.

[151] Soloviev, V., McGinty, J., Tahir, K., Neil, M., Sardini, A., Hajnal, J., Arridge, S. and PM, F. (2007). Fluorescence lifetime tomography of live cells expressing enhanced green fluorescent protein embedded in a scattering medium exhibiting background autofluorescence. *Optics Letters*, volume 32(4), pp. 2034–2036.

[152] Srinivasan, S., Pogue, B. W., Jiang, S., Dehghani, H. and Paulsen, K. D. (2005). Spectrally constrained chromophore and scattering near-infrared tomography provides quantitative and robust reconstruction. *Applied optics*, volume 44, number 10, pp. 1858–1869.

[153] Suhling, K., French, P. M. and Phillips, D. (2005). Time-resolved fluorescence microscopy. *Photochemical & Photobiological Sciences*, volume 4, number 1, pp. 13–22.

[154] Szabo, T. L. (2004). *Diagnostic ultrasound imaging: Inside out*. Academic Press.

[155] Tanifuji, T. and Hijikata, M. (2002). Finite difference time domain (FDTD) analysis of optical pulse responses in biological tissues for spectroscopic diffused optical tomography. *IEEE Transactions on Medical Imaging*, volume 21(2), pp. 181–184.

[156] Tarvainen, T., Vauhkonen, M. and Arridge, S. R. (2008). Gauss-Newton reconstruction method for optical tomography using the finite element solution of the radiative transfer equation. *Journal of Quantitative Spectroscopy & Radiative Transfer*, volume 109, pp. 2767–2778.

[157] Thom, A. and Apelt, C. J. (1961). *Field computations in engineering & physics*. Van Nostrand.

[158] Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. W.H. Winston.

[159] van der Walt, S., Colbert, S. C. and Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science Engineering*, volume 13, number 2, pp. 22–30.

[160] van Ginneken, B., Schaefer-Prokop, C. M. and Prokop, M. (2011). Computer-aided diagnosis: How to move from the laboratory to the clinic. *Radiology*, volume 261, number 3, pp. 719–732.

[161] Venugopal, V., Chen, J., Barroso, M. and Intes, X. (2012). Quantitative tomographic imaging of intermolecular FRET in small animals. *Biomedical Optics Express*, volume 3, number 12, pp. 3161–3175.

[162] Wang, G. (2016). A perspective on deep imaging. *IEEE Access*, volume 4, pp. 8914–8924.

[163] Wang, L., Jacques, S. L. and Zheng, L. (1995). MCML-Monte Carlo modeling of light transport in multi-layered tissues. *Computer methods and programs in biomedicine*, volume 47, number 2, pp. 131–146.

[164] Wang, L. and Wu, H. (2007). *Biomedical Optics - Principles and Imaging*. Wiley-Interscience, A John Wiley and sons, Inc.

[165] Wang, X., Pogue, B. W., Jiang, S., Dehghani, H., Song, X., Srinivasan, S., Brooksby, B. A., Paulsen, K. D., Kogel, C., Poplack, S. P. *et al.* (2006). Image reconstruction of effective mie scattering parameters of breast tissue in vivo with near-infrared tomography. *Journal of biomedical optics*, volume 11, number 4, pp. 041106–041106.

[166] Weissleder, R. and Ntziachristos, V. (2003). Shedding light onto live molecular targets. *Nature Medicine*, volume 9, number 1, pp. 123–128.

[167] Woodbury, K. (2002). *Inverse Engineering Handbook*. Handbook Series for Mechanical Engineering, CRC Press.

[168] Yao, Y. and Wang, Y. (1997). Frequency-domain optical imaging of absorption and scattering distributions by a Born iterative method. *Journal of the Optical Society of America*, volume 14, pp. 325–342.

[169] Young, T., Hazarika, D., Poria, S. and Cambria, E. (2017). Recent trends in deep learning based natural language processing. *arXiv preprint arXiv:1708.02709*.

[170] Yuan, Y. (2000). A review of trust region algorithms for optimization. *ICIAM*, volume 99, pp. 271–282.

[171] Zhu, B., Liu, J. Z., Rosen, B. R. and Rosen, M. S. (2017). Image reconstruction by domain transform manifold learning. *arXiv preprint arXiv:1704.08841*.

[172] Zhu, C., Byrd, R. H., Lu, P. and Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, volume 23, number 4, pp. 550–560.